DTIC FILE COPY

②

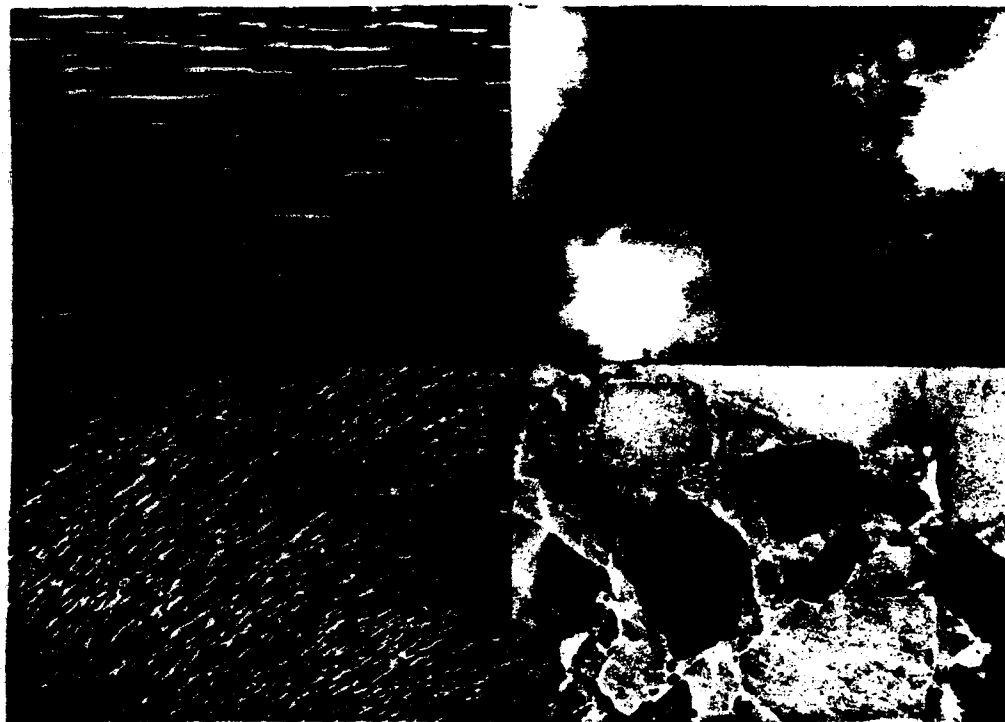# Inferences from Images:
(Final Report 1984 – 1987)

Investigators:
617-253-5776
ARPANET: WHIT@EGO.MIT.EDU
Contract No. N 00014-84-K-0650

Whitman Richards E10-120
Shimon Ullman NE43-771
Massachusetts Institute of Technology
Cambridge, Mass. 02139

DTIC
ELECTE
AUG 0 1 1989
D

ABSTRACT: How can a visual system make reliable inferences about surface properties? To answer such questions, we use computer graphics to generate a realistic rendition of a surface, such as water, clouds or rock. Such a rendition demonstrates that all the relevent psychophysical parameters have been indentified. Given these parameters, or rather their image correlates, we then proceed to show how the surface type can be inferred from the available image information (inverse optics). The surface we have studied most extensively is water, including some of its dynamical properties. Certain aspects of the motion of planar surfaces were also investigated. Our results suggest a simplification of the Cook-Torrance model for the reflectance function.

Key words: Image understanding, vision algorithms, computer graphics, biological information processing, motion, water waves, sound perception.

December 1988

89   8 - 01   017

Contents

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS
Unclassified	N.A.

2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT
N.A.	Approved for public release;
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	distribution unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)

6a NAME OF PERFORMING ORGANIZATION	6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
Mass. Institute of Technology		Office of Naval Research

6c. ADDRESS (City, State, and ZIP Code)	7b ADDRESS (City, State, and ZIP Code)
Dept. of Brain and Cognitive Sciences Cambridge, MA 02139	Arlington, VA 22217-5000

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Office of Naval Research	Code 1142PS	N00014-84-K-0650

8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS			
800 N. Qunicy Street Arlington, VA 22217-5000	PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
	61153N 42	RR04209	RR0420901	4424 188

11 TITLE (Include Security Classification)

Inferences from Images

12 PERSONAL AUTHOR(S)
Whitman Richards and Shimon Ullman

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 84/07/01 TO 87/06/30	December 1988	102

16 SUPPLEMENTARY NOTATION

17	COSATI CODES		18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Vision; visual perception; motion; material type; water; computer graphics; machine vision; short-long wave interaction; acoustics; sound perception; reflectance function.

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

How can a visual system make reliable inferences about surface properties? To answer such questions, we use computer graphics to generate a realistic rendition of a surface, such as water, clouds or rock. Such a rendition demonstrates that all the relevent psychophysical parameters have been indentified. Given these parameters, or rather their image correlates, we then proceed to show how the surface type can be inferred from the available image information (inverse optics). The surface we have studied most extensively is water, including some of its dynamical properties. Certain aspects of the motion of planar surfaces were also investigated. Our results suggest a simplification of the Cook-Torrance model for the reflectance function.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
H. Hawkins	(202) 696-4323	Code 1142PS

DD Form 1473, JUN 86   Previous editions are obsolete   SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

Inferences from Images[1]

1.0 Introduction

The goal of most researchers studying "how to see" is object recognition. The approach includes separating a scene into objects, describing these objects in terms of some model such as generalized cones (Binford, 1971; Marr,1982), and then matching the result to a canonical description in memory (see Ballard & Brown, 1982). Surprisingly, the most progress has been made in the last two steps, whereas the initial first step of identifying objects in a scene is still largely unsolved in computer vision and is not understood at all for biological systems, except in special cases where an isolated object moves against a stationary background, or is distinguished in three-dimensions with no occlusions present.

Our work differs from most in vogue today by suggesting that a scene be separated into objects not by edge extraction followed by contour grouping, but rather by recovering regions of differing material qualities. This approach was attempted early in scene analysis (Ballard & Brown,1982), but then was based only on simple image features such as color and texture. Although a useful beginning, a more reliable method is to use the image features to infer material properties of the world (e.g. water, cloud, tree, grass, wood, asphalt, etc.) and to break up the scene into its "stuff". Such world-based properties are stable and survive changes in illumination, viewpoint, shading, surface orientation, etc., which confounded earlier image-based attempts at scene analysis (Heeger & Pentland 1987; Kube & Pentland, 1986).

By identifying the "stuff" a region is made of, we also gain the additional benefit of knowing something about what object that region might correspond to. For example, hair implies an animal, feathers . `rd, grass a surface which is part of the terrain (Richards, 1982; Richards & Boc``k 1988).

Our major findings are :

- Useful material descriptions must include both shape and reflectance predicates. (In many cases, dynamics is also needed.) Shape topology is far more important than precise metrics.

- The Cook-Torrance reflectance model can be simplified to a three-dimensional system for purposes of rendition and image understanding.

- The use of Graphics Psychophysics (i.e. image analysis by generating synthetic images) is a very powerful but time-expensive tool for image understanding.

[1] The frontispiece is a composite of computer-generated surfaces and pictures taken from Brodatz (1966). The textures for all these surfaces can be modeled by fractal processes.

|	
α	Angle between N and H
θ	Angle between L and H or V and H
λ	Wavelength
D	Facet slope distribution function
d	Fraction of reflection that is diffuse
$d\omega_i$	Solid angle of a beam of incident light
E_i	Energy of the incident light
F	Reflectance of a perfectly smooth surface
f	Unblocked fraction of the hemisphere
G	Geometrical attenuation factor
H	Unit angular bisector of V and L
I_i	Average intensity of the incident light
I_{ia}	Intensity of the incident ambient light
I_r	Intensity of the reflected light
I_{ra}	Intensity of the reflected ambient light
k	Extinction coefficient
L	Unit vector in the direction of a light
m	Root mean square slope of facets
N	Unit surface normal
n	Index of refraction
R_a	Ambient reflectance
R	Total bidirectional reflectance
R_d	Diffuse bidirectional reflectance
R_s	Specular bidirectional reflectance
s	Fraction of reflectance that is specular
V	Unit vector in direction of the viewer
w	Relative weight of a facet slope

The geometry of reflection
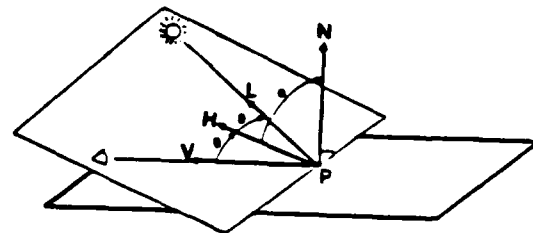
Figure 1 Parameters of the Cook & Torrance (1982) model for surface reflectance.

- We present the first study in computational acoustics showing how material properties can be recovered from sounds.

- More work is needed to understand the useful image correlates of the motion of non-rigid objects, especially those materials without surface markers such as rubber or water.

2.0 The Problem

Recovering material properties from image information is difficult because the available signal, namely the image intensity, is a function of several factors which are confounded. For example, the observed intensity (flux) I of a point on the retina depends upon (1) the spectral reflectance (albedo) of the surface, $(\rho(\lambda))$, (2) the strength and spectral composition of the illuminant $E(\lambda)$, (3) the geometry and structure of the three-dimensional surface $N(\sigma, \tau)$ relative to the light source direction L and the viewing position V, and finally (4) the reflecting properties of the material R. In particular,

$$I(x,y,\lambda) = \rho(\lambda) * E(\lambda) * (N \cdot L) * R \qquad (1)$$

Thus, if one wishes to "know" what kind of pigment $\rho(\lambda)$ makes up the surface (chlorophyll, xanthophyll, flavanoid, ferrite, etc.) the observer must somehow discount the illuminant E (including shadows), the surface orientation N, the illuminant direction L, his particular viewing position V, and any specular component of the surface. (The latter component is embedded in R, whose complete description requires many of the additional parameters defined in the table of Figure 1.) Obviously with so many unknowns, any single intensity value is insufficient (Horn, 1977, 1987). Rather, several different regions in the image of an object must be examined for evidence that will disambiguate among all the possible origins of the image intensity values (Richards, 1988). For example, to tell whether an urn is made of brass, or rubber, one must "inspect it" and integrate information from the highlight, the occluding boundary at two or more locations, the intensities arising from shaded regions, etc.

Considering the complexity of the interrelated unknowns, the problem of recovering the material property from image intensities seems impossible. Yet observers can indeed distinguish between materials such as metal, plastic, rubber, ceramic, etc., even when all other factors such as object shape and illumination are constant (see illustrations of Cook & Torrance, 1982; Richards & Ullman, 1987). On what basis are these inferences made ?

3.0 The Approach: Graphics Psychophysics

Obviously an observer can not recover all the physical parameters of a surface. Only some of these are relevant to the inference process. Our method for identifying these relevant parameters is simple: we use computer graphics to generate an image that "looks like" the real surface.

Figure 2 illustrates in more detail the role "analysis-by-synthesis" can play in Image Understanding. When a scene is imaged upon our retina, it becomes a pattern of receptor activities, as shown in the lower right of the figure. This is what we are given, namely a 2D array of numbers. Our task is to infer the structure and stuff of the world, as shown by the Inference arrow on the upper right. We face two problems. First, we need to know the numbers in the array. Second, we need to know how the numbers "got there".

For any given scene, finding the numbers in the image array is trivial. Simply take a picture with a vidicon and read the numbers out of frame buffer. We can check to be sure these numbers are correct by displaying the contents of the buffer to see if the 2D display looks like the real 3D scene. In terms of Figure 2, the
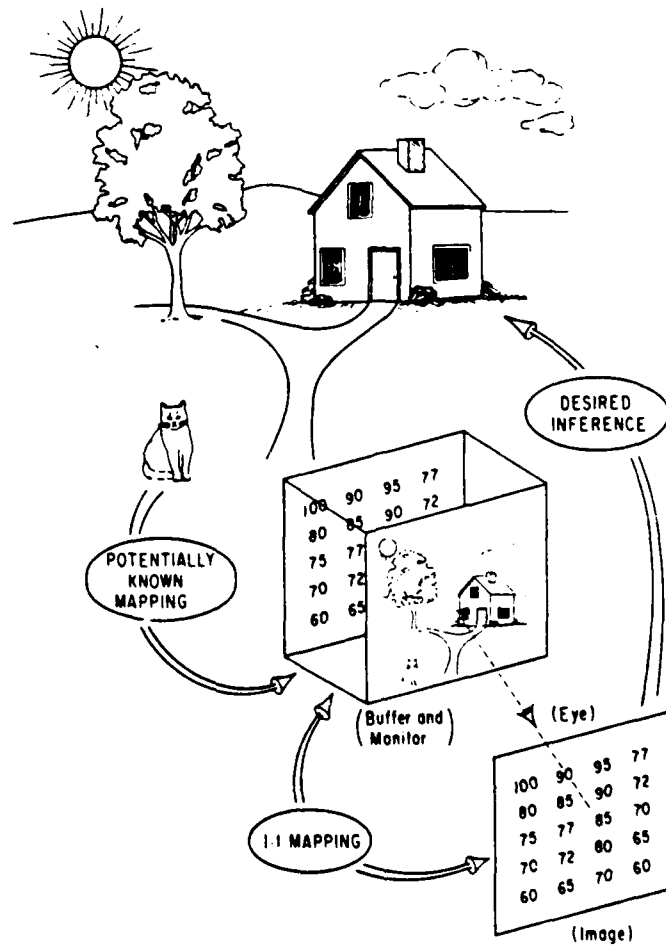
Figure 2 The role of computer graphics.

pattern of numbers in the display's buffer then matches that of the image of the scene (or at least there is a one-to-one mapping).

The tough problem is the second one, determining why the 3D surfaces in the world led to the particular array of numbers. Without knowledge of this mapping, the inference problem can't be solved, for we have not specified the function we wish to invert. In terms of our figure, we need to know how to load the frame buffer to duplicate the real 3D scene. This is a problem in computer graphics. The generation of a realistic-looking cloud or mountain forces the programmer to model the world, to make explicit the optical properties of a surface that are psychophysically relevant (Blinn, 1977, 1982). The more realistic the display, the closer the programmer's model captures the relevant properties and geometry of the surface. Specifically, we know the origin or cause of each image intensity element

$I(x, y, \lambda, t)$ on the retina of our eye, for we have specified function explicitly in our program. Hence the mapping from the world to the image plane is known. It then becomes apparent what external physical properties underly the observed property of "metal" or "rubber", as in the Cook & Torrance reflectance model (1982), or the rough surface of a mountain, as in the fractal models of Carpenter (1980), Voss (1981) and especially Pentland (1984). Computer graphics thus provides us with the opportunity to complete our understanding of the mapping of the world into the image. Once understood, possible inverse mappings and constraints can be intelligently explored. Inferences from images can proceed.

4.0 Inferring Material Types

In the sections to follow, we illustrate our Graphics Psychophysics approach with two vision examples, inferring "wood" and "water". Appendix II also includes an analogous study of how a natural property might be recovered using acoustic information.

4.1 Why Wood Looks Like Wood (D. Honig)

1.0 Introduction

The texture and pattern of a surface provides a major source of visual information about the material or "stuff" we are looking at. Once the type of material is classified, a host of other facts can be accessed, such as the expected rigidity of the surface, its hardness, its density—or other properties that may guide us in object recognition or manipulation. For the human observer, the classification of materials using optical information is almost automatic and consequently seems trivial. Do we simply store a vast collection of textures typical of surfaces of interest, learning by example the proper associations? Or, alternatively, do we first infer from the optical array the type of process which created the pattern, and then proceed with our inference? Here, we take the latter approach: how can the process which created a visible surface structure be inferred? We illustrate this approach using the pattern created by wood grain.

2.0 The Method

To understand why wood looks like wood, we begin by creating a rendition of wood, using computer graphics. This rendition is not just a texture mapping or "painting" of a pattern onto a surface, but is actually a three-dimensional construction based upon tree structure and how trees are cut up into lumber. Trees are a natural category of plant with certain family traits. Common structural properties of trees combine with the ways in which lumber is commonly cut to yield a pattern of textural elements that is practically unique to wood. It is these kinds of regularities that allow us to classify things. By using computer graphics to create a 3D model of the tree structure and then by cutting this model structure, we can identify the origin of each regularity typical of the wood grain pattern.
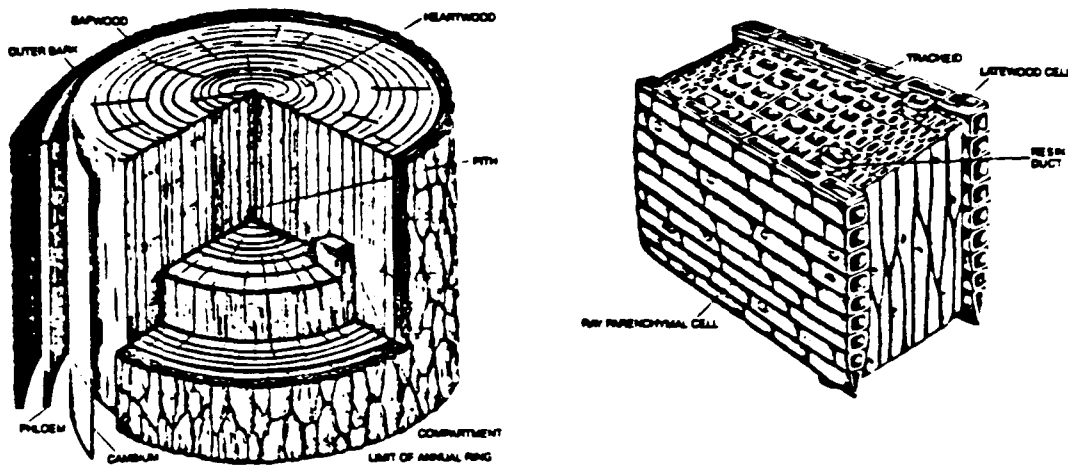
Figure 1 The structure of a tree (from Sligo, 1985).

3.0 The Processes

As previously mentioned, the pattern typical of wood grain depends both upon the natural structure of trees and upon the way trees are cut into lumber. We discuss each process separately, prior to presenting our model of these processes.

3.1 The Tree

A tree is a living process, absorbing minerals from the ground and photosynthetic energy from the sun. The trunk and its branches are cylindrical structures with three main parts: the cambium, the pores, and radial compartments (see Figure 1).

3.1.2 The Cambium

The cambium is the only live part of the tree aside from the leaves. It is a thin, vascular layer in the trunk and in all branches and roots. Bark protects the cambium on the outside. (Here we neglect the bark, which is stripped prior to cutting boards.) The cylindrical cambium shell grows outward by depositing concentric layers of xylem (wood) cells on its inner surface. Fine tubes that transport water and other minerals up the tree are interspersed throughout the woody tissue. On the outer surface of the cambium layer the phloem (inner bark) is deposited, which
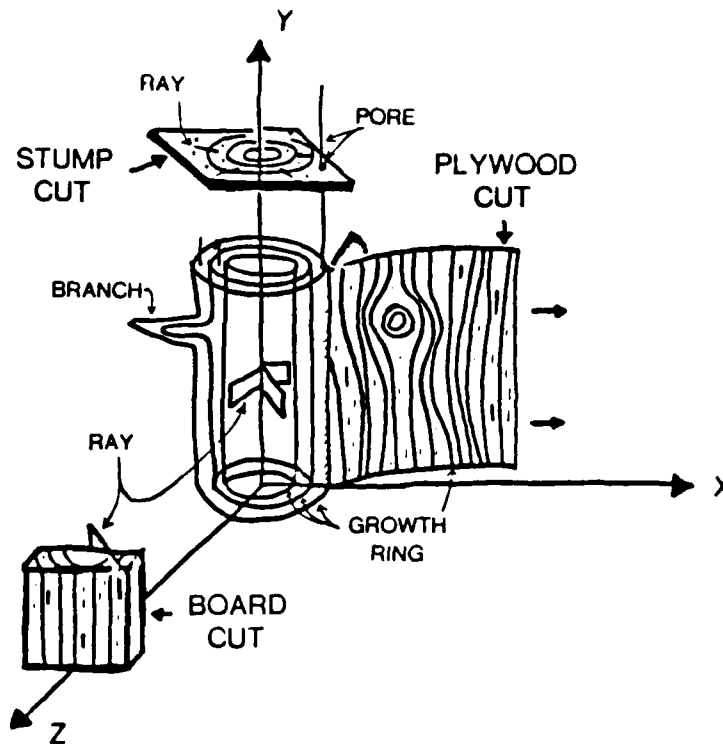
Figure 2 Typical ways in which trees are cut into lumber.

transports photosynthetic products down the tree. The outer surface of the inner bark is the bark which protects the tree.

The annual "growth rings" of a tree are formed because as the cambium grows outward, depositing new wood on its inner surface, the new tissue gets darker toward the end of the growing season. As winter approaches, the tree accretes new wood more slowly, so the natural pigmentation is denser.

3.1.3 The Pores

The tree transports various fluids through pores. These pores are simply cellulose-lined cells or fine, tubular structures oriented parallel to the axis of the trunk or branch.

3.1.4 Radial Compartments (Rays)

Trees defend themselves against infection by partitioning, walling-off damaged and infected tissue. Two of the interior walls of the partitions are radial planes of
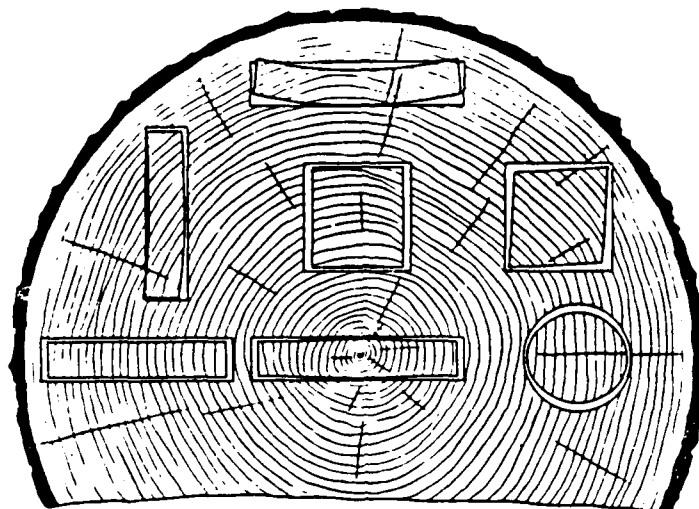
Figure 3 Several special cuts of a tree into boards. The shrinkage characteristics of lumber depend on how it is cut (from the U.S. Forest Products Laboratory).

parenchymal tissue distributed among the woody tissue. Vertically, trees stop the spread of disease by blocking the tubes (pores) passing through the site of injury. Along the radius, trees wall-off infection with the annual growth rings, whose high density serves as an effective barrier. The compartments are most visible in cross section, for wood will often crack along the radial walls of compartments.

3.2 Lumbering

When trees are cut into boards, there are typically only three major types of cuts: the "stump" cut; the "board" cut; and the "plywood" cut (see Figure 2).

Let the axis of the trunk or branch be oriented in the Y-direction. A "stump" cut is simply a plane whose normal vector is parallel to Y. The typical "board" cut is a plane parallel to Y, i.e. parallel to the axis of the branch. These cuts can have rather complicated cross-sections, as illustrated in Figure 3. The "plyboard" cut is a cylindrical shell or "stripping" parallel to the cylindrical structure of the trunk or branch. Finally, there are several "odd cuts", like spheres or dowels. Most of our renditions used board cuts, which generate the most common wood grain patterns.
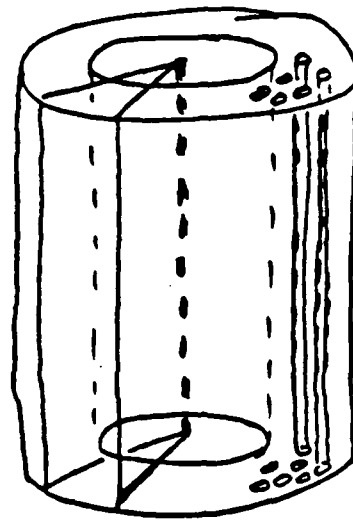
RINGS, RAYS, PORES

Figure 4 Simplified model of the three major components of a tree's structure.

4.0   The Model

Our idealized tree has three structures: a series of cylindrical shells (cambium), fine tubular vessels running parallel to the cylindrical axis (pores) and radial compartments (rays). As illustrated in Figure 4, these we identify as the "rings", "rays", and "pores".

The "rings" in the computer model were created by concentric cylinders with a fixed shell width, but having a gray scale gradient across the shell. The pores were simulated by randomly distributing tiny vertical tubes throughout the modelled tree trunk (i.e. the nested cylinders). The rays were modelled as radial vertical rectangular planes of bounded radial and vertical extent. These planes were rendered darker than the backgound material, spaced at roughly 20° intervals.

Although wood is usually light brown, some varieties like "rosewood", "cherry", or basswood can exhibit a range of pastel colors. Furthermore, if wood is stained or ṇ ṇted, it can be of almost any color. Here, we generally used 256 brightness leveɪ of the same color. One part red to one-third green to one fifth blue in the C ˙ ˑ monitors produced a realistic wood-brown.
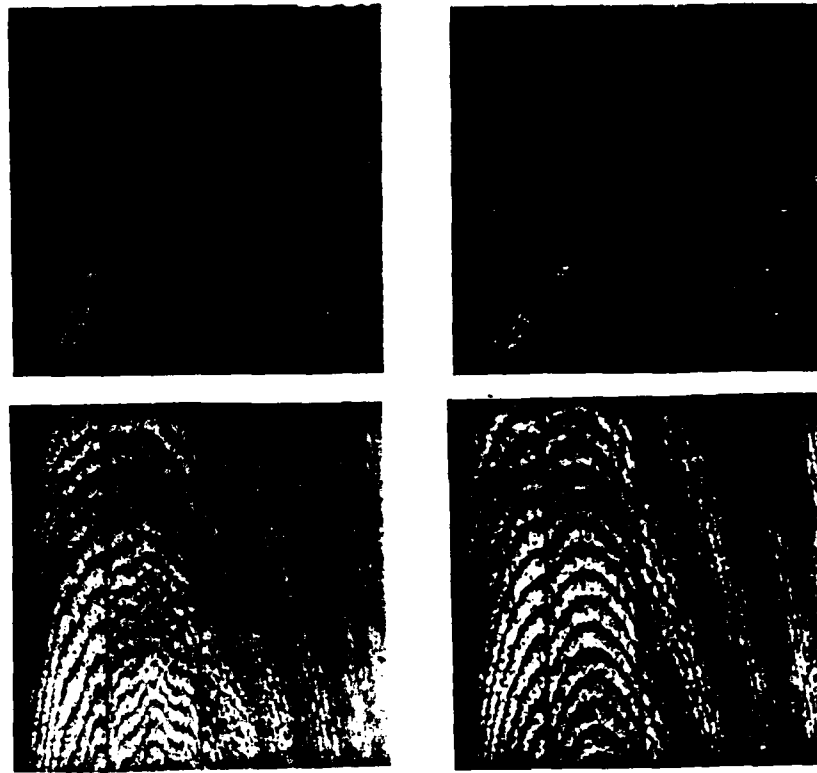
Figure 5 Renditions of "wood".

5.0 The Renditions (Instant Psychophysics)

Each rendition can be characterized by the plane of section through the tree trunk (or branch). With the Y-axis of the coordinate system aligned with the central axis of the tree, and with the Z-axis pointing to the viewer, any plane can be described by its normal vector $[x, y, z]$.

Figure 5A shows a "stump cut", $N = [.2\ 1\ .1]$, overlaid with the tapered end of a dowel, $N = [-1\ 1\ 0]$. For the stump cut, we see the rings manifest as circles, with the rays appearing as radial lines and the pores as points. We have also added a slow radial modification in albedo to model longer-term variations in density due to climate. This addition breaks up the otherwise regular geometric pattern and produces a more acceptable rendering. As an aside, we note here that the particular waveform of the annual density of wood is not very critical (i.e. the cross-sectional intensity profile of the rings). Technically the most dense (and hence the darkest) portion of the ring should be on the inner surface of each layer of the cambium. In practice, the typical observer ignores this subtlety—we have reversed the phase of the pattern with no loss in rendition quality.

The remaining figures show "board cuts" $N = [a\,0\,b]$, with a, b taking on slightly different values for each of the five panels of each figure. The most obvious feature of all these panels is the expected elliptical pattern of the rings. In Figure 1B, only the rings (plus slow climatic modulation) is illustrated. The rendition is acceptable. Adding the rays greatly improves the rendition (Figure 1C). However, breaking up the regular elliptical pattern by modulating the cut produces the most realistic wood grain. Note that an undulating plane of section is equivalent to a planar section of a tree which is not strictly a nested stack of regular cylinders. The observation that the presence of irregularities is a necessary part of a good rendition is important. It means that the inference "wood" requires some evidence that a strictly regular, machine-like process is not solely responsible for the observed structure.

In sum, our renditions suggest that the geometrical model consisting of nested cylinders, radial compartments, and pores (which create a fine-grained "peppery" texture) is acceptable provided that two kinds of perturbations are included in the strictly geometric model: 1) longer-term climatic variations in pigmentation density and 2) allowance for undulations in the cylindrical axes and diameter.

6.0 The Inference Process

A wooden surface typically exhibits a "grain direction", either from the rays or from the pores or both. For perfect planar cuts, this grain direction will range from radial (stump cut) to parallel (board cut). The major axes of the ellipses of the growth rings must align with this grain direction, all intersecting together at the center of the tree. In addition, the spacing between the elliptical rings must be the same along any given ray (or grain direction). Furthermore, the rings must exhibit an albedo variation whose function is normalized for the different separations between the rings for different rays. This is the strictly geometric model. In addition, some evidence must be presented that the pattern is not simply a regular geometric structure. Slow albedo modulation and perturbations in the shape of the ellipses along a given ray seem to provide convincing evidence of a natural growth process.[1]

7.0 False Targets

One simple approach to enumerating the important false targets is to recognize that the strictly geometric model results in an image of concentric ellipses with equal radial increments from one to the next ellipse. We now have four free parameters as this plane is moved to create a nesting of surfaces in 3-space: the ratio of major

---

[1]Note: saw cuts may also create density variations parallel to the edge of the boards.

to minor axis, the length of the major axes, and two degrees of freedom in the direction of movement of the common axis (i.e. center of the ellipsis).

An onion falls into this class of possible generating surfaces. In this case, the major and minor axes of the ellipse are equal (circles), and the plane of nested circles is moved in a direction perpendicular to the original plane of the circles, and the major axis (radius) is decremented as we move away from the initializing plane, yielding a set of nested circles (or ellipses if we wish). A nesting of cones also could yield a false target.

To exclude these possibilities, we may require further evidence, such as radial "cracks" or albedo variations in the ring pattern. But the key reason for the success of the inference "wood" from the image slice is that in the natural world, structures created from nested cones or ellipses are very, very rare—indeed it is difficult to give a natural example of such a structure! A nesting of spheres is possible, such as the onion, but here the potential for confusion is only with the less common "stump cut". Fortunately, Nature creates structures according to rules which tend to group similar classes together, creating "modes" of particular structures rather than a continuum of all possible functions (Bobick, 1987). The cylindrical construction of plants and trees is an example of such a "mode" which permits successful inferences from the pattern we recognize as "wood".

8.0 Summary

One might ask, was computer graphics necessary to determine the characteristic features of wood? Couldn't we have just sketched the projections of the main components of the tree projected onto the image plane, just as an artist might do?

Certainly an artist's rendition of any scene is informative and helpful. But can the artist tell you exactly what each brush stroke represents in terms of a world property? Without such knowledge, the inference from image to world is not possible. Furthermore, the artist is depicting his internal model of the material, which is not necessarily an accurate world model—certainly not a physical model. Often, certain features are exaggerated for emphasis, and usually only common views are depicted. The graphics model, on the other hand, is not limited to special views and exaggeration must be made explicit. Hence the graphics model can be used to explore unusual viewing or illumination conditions to determine the actual character of the observer's internal model. Such power is needed if the observer's representation and inference process is to be understood fully. Further examples of this point follow in subsequent sections, where unusual viewing conditions of a "correct" physical model lead to mis-perceptions because the observer's model assumes particular viewing and illumination conventions.

9.0 References

Sligo, A.L. (1985) Compartmentalization of decay in trees. *Sci. Amer*, **252**, April.

Bobick, A.F. (1987) Natural object categorization. *M.I.T. Artificial Intelligence Laboratory Technical Report 1001* (revision of thesis submitted to the Department of Brain and Cognitive Sciences, M.I.T.).

4.2 Inferring "Water" from Images (T.J. Kung)

1.0 Introduction

Water waves are recognizable from a black and white photo. What is the image information that allows us to make this inference? In order to answer this question, we use computer graphics to simulate water. Once a perceptually deceiving simulation of water is obtained, then the study of the inference process can proceed. This report presents a model for generating near-natural images of water waves.

Computer generated water waves have been created previously for flight simulators. However, these images did not look very realistic, probably because very simple models were used. For instance, Max (1981) presented a ray-tracing procedure in which ocean waves and islands are rendered. Although he was able to render good reflection images off waves, the ocean waves do not look like real waves. Ogden (1985) generated a real-time animation sequence of water by using the Burt Pyramid (1983), but again her model was not based on the physics of water. Furthermore, because the method essentially generated fractals in the image plane, there was no real description of the 3D surface. Here, our objective is to create a 3D surface of water undergoing wave motion, and to be able to view this surface from any arbitrary angle. Complicated effects like the reflection of an object off the water surface are not considered in this paper.

We divide the problem into three parts: (1) reflection, (2) surface shape, and (3) dynamics. In interpreting images, basically we are trying to infer the surface properties and shape from the image intensities. As a first step in understanding this inference process, we must understand what properties make water characteristically different from other surfaces. If we knew these properties and how to describe them mathematically we should be able to create a realistic image. Among the surface properties, we are only interested in the optically related properties, i.e. the reflectance function. A dynamic view yields a considerably stronger impression, therefore a brief discussion of wave dynamics is included also.

2.0 Reflectance Function

Because water is a transparent material, the intensity and color of the visible light emerging from the sea depends on

(i) *The illumination.* The elevation of the sun and the sky and cloud cover determine the quantity and spectrum of downwelling incident light.
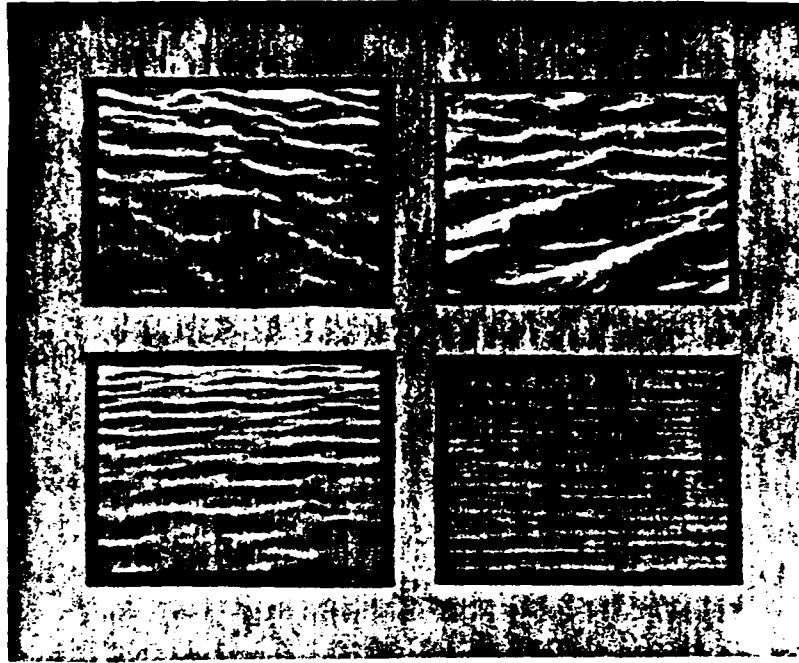
Figure 1 Synthetic water images are shown in the two left-hand panels. The upper right panel is an artist's rendition. A photographic image of water is shown in the lower right. Most people prefer one of the synthetic images, demonstrating that our model is capturing the psychophysically relevant information.

(ii) *The optical properties of the sea-water itself.* The light which penetrates into the sea is spectrally modified by absorption inside the medium before being partly backscattered toward the atmosphere. These absorption and backscattering processes have their origin in the water molecules, and also in all the other substances present in the sea as dissolved or particulate matter.

People can recognize water waves from a black and white photo, therefore the most important thing in perceiving water waves is the relative intensity of emerging light from different places of the water surface, not the details of spectrum and quantity. This fact makes it possible to simplify the reflectance function without considering the functional dependency on the wavelength of lights.

Cook & Torrance (1982) proposed a reflectance model for computer graphics as follows:

$$I = R_{amb}E_{amb} + \sum E_{dir}(\mathbf{N} \cdot \mathbf{L})(dR_{mat} + sR_{spec}) \tag{1}$$

where E is the intensity of light source, $\mathbf{N}$ is surface normal unit vector, and $\mathbf{L}$ is the light direction of a point light source. The point light sources are linearly
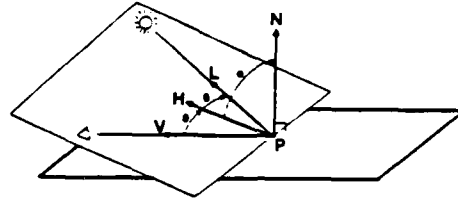
Figure 2 The geometry of reflection.

combined with the ambient light source. Each point light source is further divided into a specular component and matte component. The d and s in equation (1) are the fractions of reflectance that are matte and specular. The specular component is given by:

$$R_{spec} = \frac{F}{\pi} \frac{DG}{(N \cdot L)(L \cdot V)} \tag{2}$$

where F is a Fresnel term which describes how light is reflected from each smooth microfacet. It is a function of incidence angle and wave length. The geometrical attenuation factor G accounts for the shadowing and masking of one facet by another. The facet slope distribution function D represents the fraction of the facets that are in the direction H (Figure 2).

In rendering water, we use Cook & Torrance's model with some modifications. As discussed before, water backscatters part of the incident light. Although the backscattered light is a function of the incident light spectrum, optical properties of the water itself, and the elevation of viewing position, we assume these conditions are fixed in a limited area and within a short period of time. Therefore the backscattered light is a constant and is uniform over the water surface.

Figure 3 shows the value of Fresnel term for various indexes of refraction n. The Fresnel term is significant only when the incident angle is greater than 70 degrees. For water the index of refraction equals 1.32. Based on this property, we make two simplifications. First, we neglect the matte term because a water surface only reflects light in very limited directions. Second, we neglect the point light source-sun. By doing so we eliminate the special cases when it is dawn and when it is dusk, and also scintillation effects which are easy to mimic.

The ambient light source (sky) is considered as a hemi-sphere light source. No matter where the viewing position is, it is always possible to see some lights from somewhere in the dome directly reflected off the water to the eyes. When the viewing angle is less than 70 degrees, so is the incident angle, and the reflected
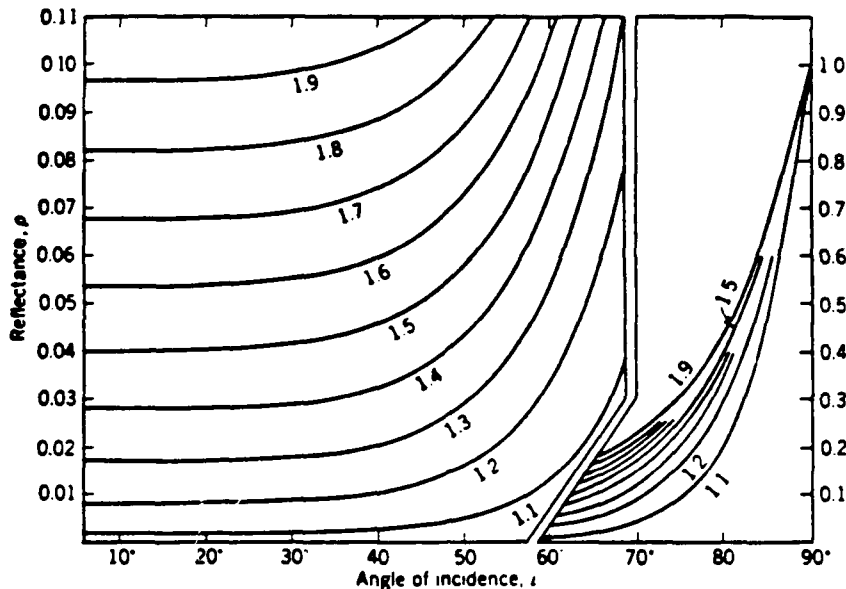
Figure 3 Fresnel reflectance as function of incident angle. The parameter on the curves is the index of refraction, n. Note the change in scale for the right portion of the graph.

light is much less than when the viewing angle is greater than 70 degrees. In other words, the intensity of reflected light from the water surface is a function solely of the angle between the viewing direction and the surface normal of the facet.

As a conclusion of above discussion, our reflectance model for water waves is

$$I_{total} = I_{b_s} + I_{amb} \int_{dome} \frac{FDG}{(N \cdot L)(L \cdot V)} d\omega \tag{3}$$

where I_{b_s} is the backscattered light intensity and is a constant. The integration term is for the ambient light source. Since this term is a function of the angle between N and V only, we calculate these values numerically for different angles and store them in a table. In rendering the images, we then just find the value of this integration term from that table. In order to simplify the integral work, the Gaussian model is used for the facet slope distribution function D

$$D = \exp\left(-(\alpha/m)^2\right) \tag{4}$$

where α is the angle between N and H (Figure 2). 0.8 is used for m. With this value D is small when α is greater than 20 degrees and then the integration can be neglected. In other words, the integrated region can be reduced from a hemi-sphere to a cone of 20 degr e radius. The expression for G is

$$G = \min\left[1, \frac{2(\mathbf{N} \cdot \mathbf{H})(\mathbf{N} \cdot \mathbf{V})}{(\mathbf{V} \cdot \mathbf{H})}, \frac{2(\mathbf{N} \cdot \mathbf{H})(\mathbf{N} \cdot \mathbf{L})}{(\mathbf{V} \cdot \mathbf{H})}\right]. \tag{5}$$

The cone is further divided into three areas based on which term in Eq. (5) is important. The full expression of F used in the integration term is

$$F = \frac{1}{2} \frac{(g-c)^2}{(g+c)^2} \left[1 + \frac{[c(g+c)-1]^2}{[c(g+c)+1]^2}\right] \tag{6}$$

where

$$c = \mathbf{V} \cdot \mathbf{H}$$
$$g^2 = n^2 + c^2 - 1.$$

For water, n is 1.32. The ratio of I_{bs}/I_{amb} was set to 1/3 in our simulation.

3.0 Surface Shape

Traditionally, sinusoidal functions are used to model ocean waves (Bigelow & Edmondson, 1952; Gross, 1972). There is a strong theoretical basis to assume that the ocean surface is formed by sums of long crested sinusoids:

$$w(x,y) = \sum\left[C + B * \cos\left(\frac{2\pi}{\lambda_x}x + \frac{2\pi}{\lambda_y}y + \phi\right)\right] \tag{7}$$

Since this model requires the sum of a large number of wave forms to obtain a realistic picture, it is not practical for image generation. Schachter (1980) proposed a narrow-band noise model to overcome this problem. In this model, it is assumed that envelopes and phase shifts are slowly varying:

$$w(x) = a(x) * \cos(\mu x + \phi(x)) \tag{8}$$

Although this model has the advantage of being able to simulate different textures, the results show that the water generated by this model does not have the shape randomness that real water does.

We propose another model which is a summation of squared of sine waves with slowly varying wave length:

$$w(x,y) = \sum\left[A * \left(1 + \sin\left(\frac{2\pi}{\lambda(x,y)}(x\cos\beta + y\sin\beta) + \phi\right)\right)\right]^2 \tag{9}$$

Here we assume waves traveling in the x, y plane and β is the angle between x-axis and the traveling wave direction. The ideal wave form is sinusoidal but a lot of factors can distort waves from their ideal forms, for instance winds, nonlinearity, and nonuniform water depth for shallow water waves (Bascom, 1980). Normal wave

Component Number	1	2	3	4	5
Mean wave-length ℓ	25	30	37.5	10.7	16.7
Amplitude A	0.3	0.4	0.4	0.15	0.25
Modulation Rate r	0.009	0.013	0.016	0.0	0.0
Angle (dgr) β	110	90	65	120	60

Table 1 Static wave parameters.

crests tend to go steep before collapse. Consequently, the distorted wave form is steep at the crest and flat at trough. We found it takes less computation to get the distorted wave form by using a function which has its shape close to the distorted wave form such as the sine squares than by using Fourier series (summation of sinusoidal waves). [See Appendix II.]

Because of nonlinear effects, wave trains are always modulated when traveling in a field. Instead of assuming slowly varying amplitude and phase shift as Schachter does, we assume slowly varying wave length. The amplitude is assumed to be a constant because the variation is too small to be noticed in short distances.

Nonlinear instability theory tells us that instability causes steep gravity waves to evolve into the three-dimensional spilling breaking waves. This three dimensional wave form is the normal wave form we see in an open sea. To simulate the three dimensional wave patterns, it is necessary to have at least three components in equation (9). It was found that to generate a nice looking water wave image, each component should travel in a direction about 20 to 30 degrees apart from the other. The image looks even better with one more minor component on each side of the three major ones (with amplitude and wavelength about one third of the average values of the major components).[1] In Table 1 these directions are specified by parameter β which is the angle measured counterclockwise from the horizontal axis (x-axis). After testing several functions, we decided to use the following function for the slowing varying wave length

$$\lambda(x, y) = \frac{\ell}{1 + r\sin(px + py)} \tag{10}$$

where l is the mean wave length and r is the normalized deviation of wavenumber, and p is the parameter which shows how slowly the wavelength varies. In our

---

[1] Texture psychophysics has shown that only 5 or 6 components are required to mimic very complex sums of sinusoids (Richards, 1979).

simulation p was fixed at 0.1. The actual values of the other parameters are listed in Table 1. The unit for wave length and amplitude is the pixel.

4.0 Wave Dynamics

The wave dynamics are solved from Navier-Stokes equation (Currie, 1975). With the assumptions that the Coriolis force are negligible, water is inviscid and incompressible, density is constant, and water motion is irrotational, the Navier-Stokes equation can be reduced to

$$\nabla^2 \psi = 0 \tag{11}$$

where ψ is the velocity potential. There are two boundary conditions at the air-water interface. The kinematic boundary condition says that particles at free surfaces remain on the free surface and the dynamic boundary condition says that the forces balance at free surface. One more boundary condition says that the vertical component of fluid velocity is zero on the ocean floor or bottom.

The surface tension can not be neglected from the dynamic boundary condition when the wave length is less than 7 cm. These kind of waves are just the ripples on top of the larger gravity waves. By gravity wave, we mean that the gravity force is the only restoring force for a wave system.

For a gravity wave, if the wave length λ satisfies

$$h > 0.28\lambda \tag{12}$$

where h is the ocean depth, then this wave is called a short wave or deep water wave. In this case, a wave travels with velocity

$$C = \frac{1}{2}\sqrt{g\lambda/2\pi} \tag{13}$$

Because this velocity depends on wave length, longer waves travel faster than shorter waves and the wave system is dispersive. On the other hand, if the wave length satisfies

$$h < 0.07\lambda \tag{14}$$

This wave is called a long-wave or shallow water wave. In this case, the waves travel with the same speed

$$C = \sqrt{gh} \tag{15}$$

Component Number	1	2	3	4	5
Mean wavelength l	25	30	37.5	10.7	16.7
Deep water $(h > 0.28\lambda)$	0.9ϕ	1.0ϕ	1.1ϕ	0.58ϕ	0.74ϕ
Shallow water $(h < 0.07\lambda)$	1.0ϕ	1.0ϕ	1.0ϕ	1.0ϕ	1.0ϕ

Table 2 Phase shift ϕ between frames used to generate dynamic water waves.

Therefore waves are nondispersive in shallow water.

In order to have an idea about how these two modes of waves look, we simulated traveling waves in both cases based on our static model described before. We generated 40 static pictures, being the maximum allowed by our computer memory. Each picture was phase shifted from the previous one. For deep water waves this phase shift was proportional to the square root of wavelength and for shallow water waves it was independent of wavelength. The exact values of phase shift we used are listed in Table 2. We cycled these pictures sequentially and video taped the whole procedure. It turned out that the deep water mode looks better than shallow water mode. This result is consistent with our daily life experience. Except for shore waves and turbulent streams, we hardly see a pond, river, or sea with water depth less than a quarter of the average wave length.

5.0 Conclusions

A procedure for generating near-natural water-wave images by a computer has been presented. Three factors play the dominant role in our perception of static water images. The first is the Fresnel term in the reflectance function. The second is hemispheric illumination. Although a point light source, such as the sun, was not included in our simulation, this is obviously the key factor for the speckle and scintillation effects of water (Longuet-Higgins, 1960). The shape and composition of the wave-itself was also found to be important. For dynamic water-waves, it is necessary to include non-linear terms that mimic the behavior of deep water wave systems.

6.0 Appendix I: Image Features

In the main body of this paper we used computer graphics to determine which physical properties were significant factors in creating images that looked like water.
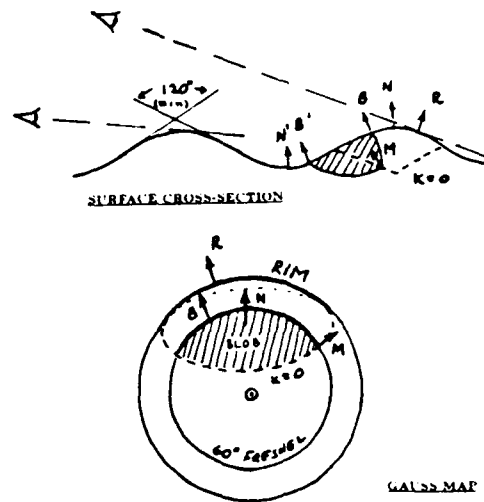
Figure 4 The top panel shows a cross-section through a surface formed by
two orthogonal sinusoids of equal wavelength and amplitude. The cutting plane
contains the line of sight and two vertical surface normals, and hence is a special
view. The cross-hatched region will appear as a darker "blob", defined by an
isoluminance contour corresponding to an emittance angle of 60°. In the lower
panel, the surface normals are mapped on the Gaussian sphere, with the viewer
direction normal to the plane of the page. The line of zero Gaussian curvature
is shown as dashed ($\kappa = 0$).

Ignoring dynamics, these were i) Fresnel reflectance, ii) hemispheric illumination,
and iii) a wave packet composed of (non-linear) sinusoids of low amplitude. Here
we proceed with the second step of our graphics approach to image understanding,
namely identifying an image feature characteristic of "water" or other fluids in
motion.

The top half of Figure 4 shows a cross-section of a water wave viewed from an
angle of about 20°. The visual ray is tangent to the wave at R, which is a point
on the "rim" of the surface (Koenderink & van Doorn, 1982). The locus of all such
rim points will be a great circle on the Gauss map. If we now place the viewer
direction normal to the page (bull's-eye), then the rim will be the outline of the
Gaussian sphere, as illustrated in the lower half of the same figure.

Consider now the isoluminance contour of 5% reflectance on the wave surface.
For a Fresnel reflectance function, this corresponds to a 60° viewing angle. We
wish to determine the general shape of this contour, and hence of the dark region
it encloses. Referring to the Gauss map, the region of 60° Fresnel reflection will be

described simply by a circle cen'ered about the bull's-eye, assuming hemispheric illumination without wavelet occlusions.

To determine which part of the 60° Fresnel circle is visible, consider the locus on the sinusoidal surface which has zero Gaussian curvature ($\kappa = 0$). For our surface waveform of crossed sinusoids, this locus can be shown to be a distorted square (or rectangle) on the Gauss map, with the square centered about N. The region contained by both the Fresnel circle and the $\kappa = 0$ square thus represents the surface orientations of the dark "blob". On the Gauss map, this shape will have cusps at the extremal views. (Obviously the exact choice of cut-off for the angle of incidence is not critical.)

If the "blob" region on the Gaussian sphere is now mapped back onto the surface, we obtain two dark regions abutting at the locus $\kappa = 0$. We wish to show that the extremal ends of this region, namely where the surface normal has orientation M, will be seen as "cusped". (A cusp on the Gauss map in itself is not a sufficient condition for a cusp on the surface.) One approach is to solve analytically for the isoluminance contour and show that the first derivatives are discontinuous here. Alternatively, we note that as we move toward M along the blob contour, for there not to be a cusp at M, the isoluminance contour must be perpendicular to the parabolic line $\kappa = 0$, because M is an extremal point of the contour. But generally this will not be the case, because the behavior of the isoluminance contour and the $\kappa = 0$ contour are controlled by independent processes, and hence generally will intersect transversally as they do on the Gauss map. The extremal end-points of the dark "blobs" of water should thus be cusped. (Close inspection of photos of water confirm this prediction.)

In the above, our analysis assumed a special viewing position and a special waveform. First, it should be obvious that the cusped property of the blob will remain unchanged if the wavelength of either one of the sine waves is increased (except in the limit as the wavelength becomes infinite). Similarly, we expect no effect of small non-linearities of wave shape. However, it is clear that viewing angle can alter the geometry of the blob drastically, for example when the surface normal N approximates the viewer direction, such as when we look almost directly down at water. In this case the tangent to $\kappa = 0$ at M may lie in the plane of the view vector, causing the cusp angle to reach $\pi/2$, thereby eliminating the discontinuity at M. The cusped feature for water is thus guaranteed only for shallow ($< 30°$) viewing angles.[2]

One final point can be made about the shape of these dark "blobs" characteristic of water. Once the amplitude of a wave is such that its bitangents intersect at

---

[2]Note that for very shallow view angles the lower portion of the dark blob of one wave will be occluded by the bright ridge of the next nearer wave, also resulting in cusped ends, or even in an extra pair of cusps, depending upon the view angle.
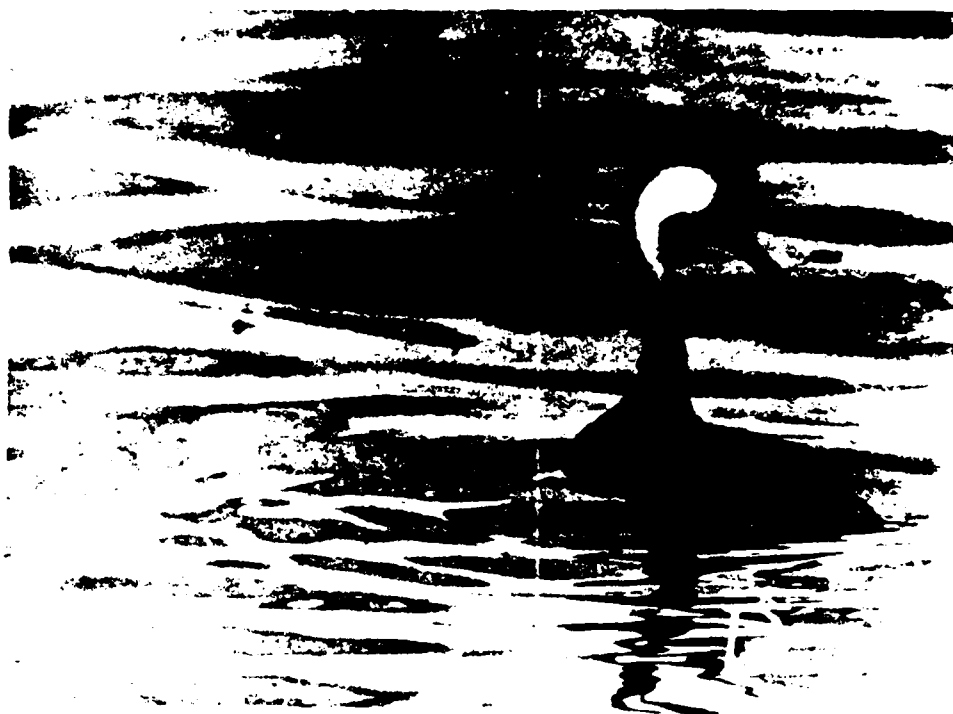
Figure 5 A "zoom" photo of a water scene, showing clearly the predicted cusps
and image features (credit for photograph – unknown).

less than 120° (see Figure 4), then the wave will crest. Because the wave silhouette
(rim) sets bounds on the shape of the blob, it too must be elongated sufficiently in
order not to violate this constraint. Our characteristic image feature for water will
thus be a very elongated, horizontal elliptical-like blob having cusped end-points at
its extremal viewing positions. It is of interest to note that Ogden (1985) created
an illusion of a water surface by convolving with image noise a mask consisting of
two such blobs of opposite sign (simply by splitting an ellipse). Similarly, properly
chosen packets of Gabor filters can also generate "water-like" textures by (thresh-
olded) convolution with noise (Daugman, 1988). Both of these masks create image
patterns with dark, elongated, cusped blobs.

7.0 Appendix II: Wave Shape

Here we illustrate the psychophysical effects of three different choices for the shape
of water waves.

    In the top row of Figure 6 the non-linearity in the shape of each wave is created
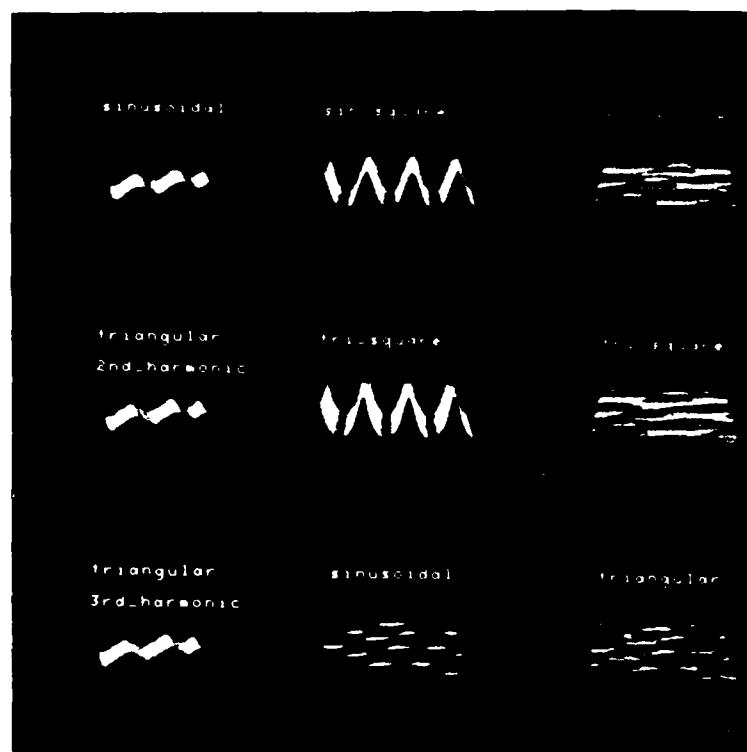simply by squaring a sinusoid (this was our final choice).

Figure 6 The psychophysical effect of the shape of the water wave upon its appearance, using Fresnel reflection and hemispheric illumination.

For comparison, the next row uses a square of a triangular wave plus its second harmonic. Note the negligble difference in the appearance of the two types of waves under fresnel reflection and hemispheric illumination (top and middle right).

Finally, the last row shows the appearance of water if only the linear sinusoid or the fundamental component of the triangular waveform are used. Neither of these renditions looks as good as the non-linear choices above.

8.0 References

Bascom, W. (1980) *Waves and Beaches*. New York: Anchor Press/Doubleday.

Bigelow, H.B. & Edmondson, W.T. (1952) Wind waves at sea, breaks and surf. . U.S. Navy Hydrographic Office, H.O. Pub. No. 202.

Burt, P.J. & Adelson, E.H. (1983) The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, COM-31:532−540.

Cook, R.L. & Torrance, K.E. (1982) A reflectance model for computer graphics. *ACM Trans. on Graphics*, 1:7−24.

Currie, I.G. (1975) *Fundamental Mechanics of Fluids*. New York: McGraw-Hill.

Daugman, J.G. (1988) Complete discrete 2D Gabor transforms by neural networks for image analyses and comparison. *IEEE Trans. on Acoustics, Speech & Singnal Processing*, ASSP-36, No. 7, July.

Gross, M.G. (1972) *Oceanography*. New York: Prentice-Hall.

Koenderink, J.J. & van Doorn, A. (1982) The shape of smooth objects and the way contours end. *Perception*, 11:129–137.

Kung, T.J. & Richards, W.A. (1985) Rendering water. *Jrl. Opt. Soc. Am.*, 2(13):P29.

Lighthill, M.J. (1978) *Waves in Fluid*. Cambridge, England: Cambridge University Press.

Longuet-Higgins, M.S. (1960) Reflection and refraction at a random moving surface. *J. Opt. Soc. Am.*, 50:851–856.

Max, N.L. (1981) Vectorized procedural models for natural terrain: waves and islands in the sunset. *Computer Graphics*, 15(3):317–324.

Morel, A. (1980) In-water and remote measurements of ocean color. *Boundary-Layer Meteorology*, 18:177–201.

Ogden, J.M. (1985) Generation of fractals using the Burt Pyramid. *Optics News*, 11(9):100.

Richards, W. (1979) Quantifying sensory channels: Generalizing colorimetry to orientation and texture, touch, and tones. *Sensory Processes*, 3:207–229.

Schachter, B. (1980) Long crested wave models. *Computer Graphics and Image Processing*, 12:187–201.

Schachter, B. (1983) *Computer Image Generation*. New York:Wiley.

Van Dyke, M. (1982) *An Album of Fluid Motion*. The Parabolic Press, California.

5.0 A New Reflectivity Model

Spectral reflectance, or more properly albedo $\rho(\lambda)$, is probably the simplest, and certainly the most colorful surface parameter in the image-intensity equation. But in addition, we must also consider R—the reflectance function itself.

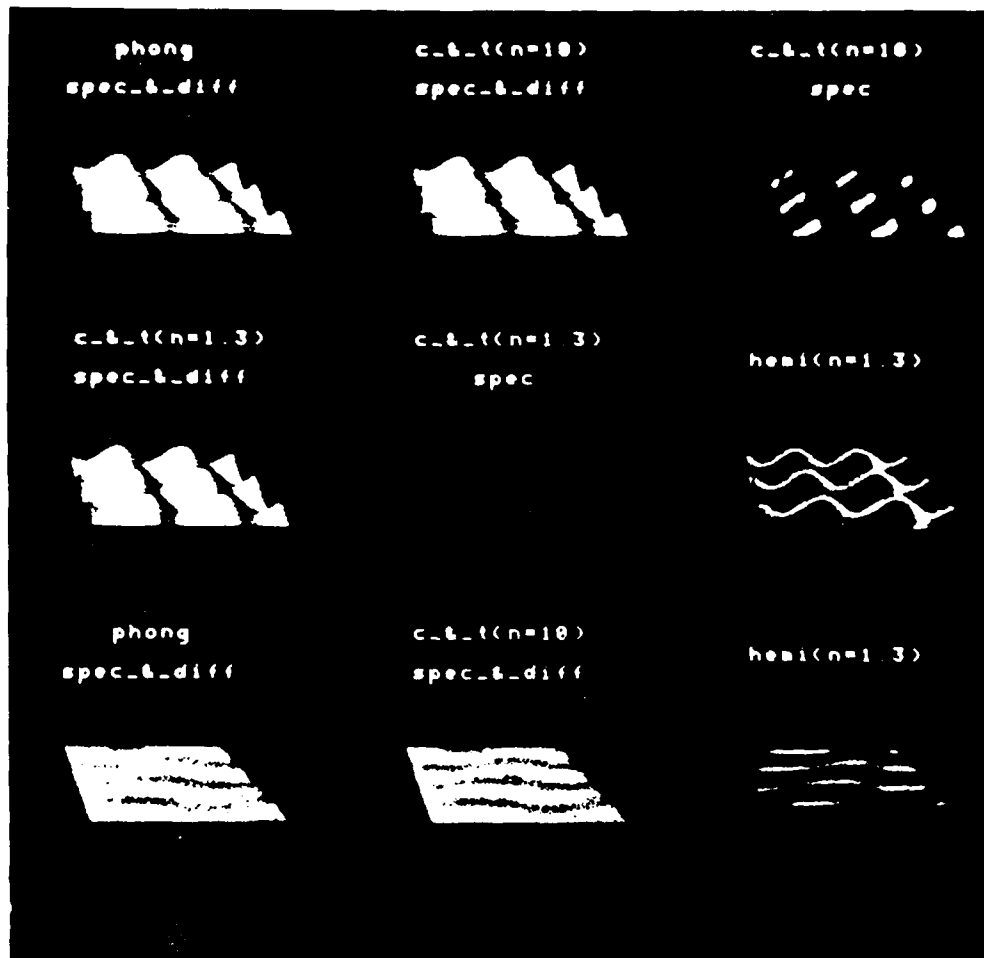$$I = \rho(\lambda) \cdot E(\lambda) \cdot (N \cdot L) \cdot R \qquad\qquad (1).$$

Figure 1 The underlying surface is two orthogonal sinusiods. The illumination and reflectance conditions can drastically alter the inferred shape and material type. Note that special conditions are required to create an image that looks like water (lower right).

Equation (1) summarizes again our basic relation between surface parameters, $\rho(\lambda)$, N; the illumination $E(\lambda)$, L, and image intensities. Cook & Torrance have elaborated this equation in detail, with special emphasis upon the reflectance parameter, R. Here, the matte versus specular aspects of the surface are collected, as well as textural effects. Following Torrance & Sparrow (1967), the textural component of the surface is considered only at the optical level, so the micro-texture itself is not visible [although see Blinn (1978) for a treatment of visible textural effects]. Although the interaction between the illumination conditions and reflectance parameters is obvious from equation (1), the full force of this interaction did not strike us until we began to model water. Furthermore, as shown in Figure 1, the *visible* surface structure, not just the *invisible* microfacet distribution, is critical to how we infer surface material type. In this figure, the underlying surface structure for all six panels in the top two rows is simply two orthogonal sinusoids at 90° to each other. The upper left panel, which uses Phong shading with some diffuse (i.e. matte) illumination, looks roughly "correct", as does the Cook & Torrance rendition using an index of refraction of 10. Note the importance of adding the matte component to a surface, for when this is taken away (upper right), the underlying surface can not be seen. Note also that the choice of index of refraction is not critical as long as there is a dominant matte (diffuse) reflectance component (compare left-most in second row with top-middle in first row). The refractive index does, of course, critically affect the specular reflectance (middle, second row versus top right, first row). Finally, rather than having just a single light source illuminating a completely specular surface (top-right or middle, second row), let the source occupy the entire hemifield (right-most in second row). The result looks more like ribbon candy or an E-M photo than crossed-sinusoids. Yet once the amplitude of the sinusoids is reduced and made typical of water, the surface indeed begins to look like a common one—namely water (lower right). Clearly there is an important interaction between the *visible* structure of the surface, its reflectance properties, and the illumination typical for that surface.

To capture this perceptual effect, we have recently proposed a new, psychophysically-based reflectance model that makes the *visible* surface roughness an explicit parameter of the representation of material type (Richards, 1988).

The basic idea is that facets on the surface may appear over a range of scales, some visible, some not, depending upon the viewing distance. What is needed is a parameter that captures this surface property. One natural choice is a fractal measure. The new representation includes the parameter h (roughness), and η (specular component), as illustrated in Figure 2. Each takes on values between 0 and 1. For example, whenever ρ, h or η is zero, then the surface will appear black, for then (i) either the albedo ρ is 0, or (ii) the surface is so rough that no light can escape to the viewer (a black body!), or (iii) the surface is a perfect mirror with
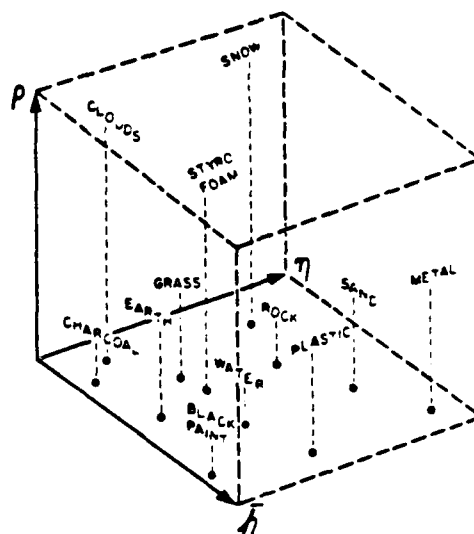
Figure 2 Representation for reflectance. Axes are: albedo, ρ; Hansdorf dimension less 1, $\bar{h}$, and a measure of the index of refraction, η, scaled to be between 0 and 1.

index of refraction $\eta = 0$, where light can not be reflected to the viewer except under very special conditions (such as retro-reflect materials). Intuitively, surfaces with a high fractal dimension (large Hausdorf-Besicovitch number) will be "rough" and hence will be dark regardless of albedo (ρ) and refractive index (η). Whereas surfaces with a low fractal dimension ($H = 1 - \bar{h}$) will be smooth, and can be either dark or light depending upon (ρ) or (η). This insight gives new meaning to "facet distribution" for *both* matte (ρ) and specular (η) reflectance. As shown by Pentland (1983, 1984, 1988) this measure can be recovered easily from image information for matte surfaces. Appendix I extends his proof to glossy surfaces. Hence a fractal facet distribution function can be estimated from images, at least, for surfaces which approximate fractals—a very large class.[2]

Thus, in our formulation for the reflectance function, all the matte properties of the surface are captured by the albedo parameter, ρ. All the specular properties thus appear in the reflectance function, R, whose two major parameters are η (index of refraction) and $\bar{h}$, which describe the surface roughness.[3] The recovery of the surface material would require estimating these three parameters from image intensities. As already noted, this is a difficult task in the general case because of the complexity of the full image intensity equation (1). However, in many cases,

---

[2] As an alternative to the fractal measure, see Lewis, 1987.

[3] Note also that $\bar{h}$ may have an orientation component, such as water does (i.e. approximately 1-D) whereas clouds are typical of a 2-D fractal function (see frontispiece). This aspect of surface texture needs further study.

conditions or surfaces exist where the image-intensity equation becomes quite simple. For example, particular locations on the surface (occluding edge, bright spot, hyperbolic region, etc.) may each provide partial information which can be assimilated. Alternately, some common surfaces like water or clouds have very special properties leading to quite simple reflectivity functions for certain very common illumination conditions such as extended, hemispheric illumination or direct sunlight (see Section 4.2).

6.0 Dynamical Properties

Most surfaces appear in motion on our retina, either because of our own movements or because the surface itself is moving. There is a large body of work analysing the image correlates of the motion of rigid objects or surfaces (see review by Ullman, and Section 5.2 and 5.3). Much less attention has been given to the motion of non-rigid objects [although see Ullman (1984), Grzywacz & Hildreth (1987) and Koenderink & van Doorn (1986)]. Even these studies have not considered the actual physics of the deformations underlying surfaces like flags waving in the breeze, or fluid motions such as clouds, oil or water. (One exception is Kajiya's (1984) study of cloud formation.) Here we begin an analysis of the dynamics of water-wave motion. Then we conclude with a more traditional analysis: the optical flow of planar surfaces. Although this study is for a rigid surface, the ambiguity is relevent to understanding the appearance of water wave motion, whose image structure looks like dark features moving in concert across the image plane. A later study of the relation between the smoothness of the (image) velocity field and surface rigidity is also relevant, reported here in Section 5.3. This second study shows that, in a structural sense, smoothness of the velocity field follows from rigidity of motion. Water waves in motion present a smooth velocity field, yet the surface is certainly non-rigid. More work is needed to clarify this ambiguity.

6.1 Non-Linear Wave Interaction Between Long and Short Waves (T-.J. Kung)

1.0 Introduction

In order to understand the appearance of waves in an image, we explore here the origin and behavior of short waves which are less than 40 cm. These waves are clearly visible not only to our eye but also to SAR (Synthetic Aperture Radar). Surprisingly, SAR images have revealed a wealth of ocean phenomena, some of which are correlated to local wind structure, long gravity waves, and large-scale circulation patterns. Such findings require a consideration of the interactions and energy transfer between wave components having both long and short wavelengths.

We begin by considering infinitesimal surface water waves, where the differential equation to be satisfied by the velocity potential is

$$\phi_{xx} + \phi_{zz} = 0 \tag{1}$$

with boundary conditions

$$\phi_z(x, o, t) = -\epsilon \frac{2\pi c}{\lambda} \cos \frac{2\pi}{\lambda}(x - ct)$$

$$\phi_t(x, o, t) + g\phi_z(x, o, t) = 0 \tag{2}$$

$$\phi_z(x, -h, t) = 0$$

where x-axis is on the undisturbed free surface, z-axis is perpendicular to the undisturbed free surface, g is acceleration of gravity, and h is the depth of water. Solving the above equations, we obtain the dispersion relations

$$\omega^2 = gk \tanh(kh) \tag{3}$$

For the case of deep water, $h \gg \lambda$ and $\tanh(kh) \approx 1$.

$$\omega^2 = gk \text{ or } c = \frac{\omega}{k} = \sqrt{g/k} \tag{4}$$

This is a dispersion wave system. The other limit is that of shallow water. $h \ll \lambda$ and $\tanh(kh) \approx kh$, therefore

$$\omega^2 = k^2 gh \text{ or } c = \frac{\omega}{k} = \sqrt{gh} \tag{5}$$

This is the non-dispersive system. The above analysis is based on the assumption that the wave amplitude a is small compared with either depth h or wave length λ whichever is smaller. Therefore the boundary conditions can be simplified. This analysis is actually an approximation of a nonlinear problem in which the nonlinear terms are, in some sense, small. Here, we want to examine how the nonlinear terms may affect a wave system through the mechanism of wave-wave interaction.

By way of background, let us consider a very elementary property of oscillation systems. A linear oscillator subjected to an infinitesimal forcing function is described by the equation

$$x_{tt} + \kappa^2 x = \epsilon e^{i\Omega t} \tag{6}$$

where κ is the natural frequency of the system and Ω the frequency of the forcing function. The response of the system from an initial state of rest is small (of order ϵ) unless κ^2 is very nearly equal to Ω^2—that is, unless there is resonance between the frequency of the forcing function and the natural frequency of the system. If the two are precisely equal, the amplitude of the oscillation grows linearly with time and becomes arbitrarily large. Indeed, the only way that the response can become large is through the phenomenon of resonance.

A very similar effect is involved in the interaction among wavetrains. Let us consider two interaction wavetrains, given to a first approximation by solutions to the linear equation.

$$a_1 \exp i(\vec{\kappa}_1 \cdot \vec{x} - \omega_1 t),$$
$$a_2 \exp i(\vec{\kappa}_2 \cdot \vec{x} - \omega_2 t),$$

$$(7)$$

where the individual wavenumbers and frequencies are related by the dispersion relation $\omega_i = \omega_i(\kappa_i)$, $i = 1, 2$. The solution to the complete nonlinear equation would be expected heuristically to be of the same general form, though the amplitude 'a' may possibly vary with time. We might seek solutions to nonlinear equation by substituting expressions of the type (7) into the small nonlinear terms. If the lowest-order nonlinearity is quadratic, this substitution leads to expressions of the type

$$\exp i\left[(\vec{\kappa}_1 \pm \vec{\kappa}_2) \cdot \vec{x} - (\omega_1 \pm \omega_2)t\right].$$

$$(8)$$

Now, these terms act as a small amplitude forcing function to the essentially linear system and provide an excitation at the wavenumbers $(\kappa_1 \pm \kappa_2)$ and frequencies $(\omega_1 \pm \omega_2)$. The response of the system to this forcing function can be expected to be small (that is, of order ϵ) unless resonance occurs—that is, unless the wave number and frequency at which the forcing is applied correspond to a (wavenumber, frequency) pair of a natural wave mode. If they do correspond, we have resonance and growth of this new component with continuous energy transfer from the two original wave components to this third one. If the initial amplitude of the third component is zero, then under resonant conditions it grows linearly until the energy drain from the other components begins to reduce their amplitudes and consequently the amplitude of the forcing function.

The above investigation is for discrete waves. A lot of research has studied the interaction phenomena between waves having widely different length scales. Since the nonlinearity is weak, it takes a long time to see zeroth order changes of the system. However, when the group velocity of a wave packet is equal to the velocity of a natural mode wave, we expect to see dramatic interchange of energy between them just like the case of discrete wave system. Let us consider the infinitesimal surface water wave again. If we have a pond with uniform depth h, then a wave with wave length λ larger than $2h$ is considered a long wave which travels at velocity $\sqrt{gh}$ (naturally defined by the system). On the other hand, we can also have a wave packet in which the wave lengths are always smaller than h. This wave packet is a deep water wave (short-wave) system corresponding to this pond therefore it travels at group velocity

$$c_g = \frac{d\omega}{d\kappa} = \frac{1}{2}\sqrt{\frac{g}{\kappa}} \text{ (in linear case)}$$

If the wave number of the short wave is right, the group velocity can be equal to $\sqrt{gh}$, the long wave velocity. This project is to study the phenomena of resonance

when the group velocity of capillary waves is equal the velocity of long waves. The method of multiple scales has been proved to be very effective in studying wave interactions and will be used in this analysis.

2.0 Theory

For the case of a progressive gravity-capillary wave moving on the free surface of a liquid of constant depth h, the undisturbed free surface corresponds to the plane $z = 0$, where z points vertically upwards, and the bottom is located at $z = -h$. The remaining Cartesian co-ordinates x, y are in the plane of the undisturbed free surface, and we choose x to point in the direction of the wave propagation. Since the fluid motion is irrotational, a velocity potential $\phi(x, y, z, t)$ satisfying Laplace's equation

$$\phi_{xx} + \phi_{yy} + \phi_{zz} = 0 \quad -h < z < \xi. \tag{12}$$

can be defined, where $\xi(x, y, t)$ denotes the position of the undulating free surface. The boundary conditions for the motion are

$$\phi_z = 0 \ \text{ at } z = -h. \tag{13}$$

$$\phi_z = \xi_t + \phi_x \xi_x + \phi_y \xi_y \ \text{ at } z = \xi \tag{14}$$

$$g\xi + \phi_t + \frac{1}{2}\left(\phi_x^2 + \phi_y^2 + \phi_z^2\right) = T\frac{\xi_{xx}(1 + \xi^2) + \xi_{yy}(1 + \xi_x^2) - 2\xi_{xy}\xi_x\xi_y}{(1 + \xi_x^2 + \xi_y^2)^{3/2}} \ \text{ at } z = \xi \tag{15}$$

Equation (13) is the boundary condition for the bottom where the normal velocity is zero. Equation (14) is the kinematic boundary condition at top surface and equation (15) is the dynamic boundary condition. The parameter T is the ratio of the surface tension coefficient to the fluid density and g is the gravitational acceleration. We suppose that initially (at $t = 0$) the surface is distorted in the manner

$$\xi(x, y, t = 0) = \epsilon\frac{i\omega}{g(1 + \tilde{T})}\left\{A(\epsilon x, \epsilon y)e^{i\kappa x} - A^* e^{-i\kappa x}\right\} \tag{16}$$

where $\tilde{T} = \kappa^2 T/g$, the asterisk denotes the complex conjugate and ϵ is a nondimensional parameter measuring the slope of the wavy surface, which has wave length $2\pi/k$. The envelope $A(\epsilon, \epsilon)$ of the surface distortion is allowed to possess a slow spatial variation and the frequency is determined uniquely by the value of κ through the dispersion relation

$$\omega = \left\{g\kappa\sigma(1 + \tilde{T})\right\}^{\frac{1}{2}}, \tag{17}$$

where $\sigma = \tanh(kh)$.

We now derive the equations describing the time evolution of A when the motion is only weakly nonlinear ($0 < \epsilon \ll 1$). On this basis we assume that (12) through (16) have a solution of the form

$$\phi = \epsilon\phi^{(1)} + \epsilon^2\phi^{(2)} + \epsilon^3\phi^{(3)} + \cdots \tag{18}$$

$$\xi = \epsilon\xi^{(1)} + \epsilon^2\xi^{(2)} + \epsilon^3\xi^{(3)} + \cdots \tag{19}$$

Also, we introduce the multiple scales

$$\xi = \epsilon(x - c_g t), \quad \eta = \epsilon y, \quad \tau = \epsilon^2 t, \quad \xi_1 = \epsilon^2(x - c_g t) \tag{20}$$

cg denotes the group velocity and is given by the relation

$$c_g = \frac{\delta\omega}{\delta\kappa} = c_p\left\{\frac{\sigma + kh(1-\sigma^2)}{2\sigma} + \frac{\tilde{T}}{1+\tilde{T}}\right\}, \quad c_P = \frac{\omega}{\kappa} \tag{21}$$

Substituting these forms into the governing set of equations, solving successively the equations resulting from repeated use of the limit process $\epsilon \to 0$, with x, y, z, t, ξ, η and τ fixed, and using the notation

$$E \equiv \exp\left\{i(\kappa x - \omega t)\right\} \tag{22}$$

one obtains the following results:

$$iA_\tau + \lambda A_{\xi\xi} + \mu A_{\eta\eta} = \nu|A|^2 A + \nu_1 A Q \tag{23}$$

$$(gh - c_g^2)Q_{\xi\xi} + ghQ_{\eta\eta} = \kappa(|A|^2)_{\eta\eta} \tag{24}$$

where

$$Q = \frac{c_g}{\kappa^2}\phi_\xi^{(0)} + \frac{c_g}{gh - c_g^2}\left\{\frac{2cp}{1+\tilde{T}} + c_g(1 - \sigma^2)\right\}|A|^2 \tag{25}$$

Equation (23) shows that the short wave satisfies the self modulated nonlinear Schrödinger equation plus an extra term—the interactions between short waves and long waves, while the modulation of long waves is totally excited by short waves as can be seen in equation (24). The coefficient ν of the cubic nonlinear term is singular when $c_g(k) = (gh)^{1/2}$. This corresponds to a long-wave/short-wave resonance in which the group velocity of the short (capillary) wave matches the phase velocity of the long (gravity) wave. Equation (23) and (24) break down under this condition and a different analysis and scaling are required.

Djordjevic & Redekopp demonstrated that the long-wave/short-wave resonant interaction occurs on a much shorter time scale than that for self-modulation. They restricted their discussion to the one-dimensional case and introduced the multiple scales

$$\xi = \epsilon^{2/3}(x - c_g t), \ \tau = \epsilon^{4/3}t, \ \xi_1 = \epsilon^{4/3}(x - c_g t), \dots \tag{26}$$

The evolution equations describing this interaction turned out to be

$$iA_\tau + \lambda A_{\xi\xi} = BA \tag{27}$$

$$B_\tau = -\alpha(|A|^2)_\xi, \tag{28}$$

where

$$B(\xi, \tau) = \delta\Phi_\xi^{(0)} \tag{29}$$

In this project we extend the analysis of Djordjevic & Redikopp to two-dimensional case. The multiple scales need to be reexamined. It is assumed that $\phi \sim (1/\mu)$ and $\mu \ll 1$ since long-wave/short-wave resonant interaction occurs on a shorter time scale. Considering equation (23), the second term on left hand side and last term on right hand side gives the relationship $A_{\xi\xi} \sim A\phi_\xi$ or $A_\xi \sim A\phi$. Therefore we can write the new streamwise direction scale in terms of old streamwise direction scale as $\xi' = \xi/\mu$. From first and second term, $A_\tau \sim A_{\xi\xi}$. The new time scale is related to the old time scale by $\tau' = \tau/\mu^2$ because $\tau' \sim (\xi')^2 = (\xi/\mu)^2 = \tau/\mu^2$. When resonance occurs, $c_g^2 = gh$, the first term of equation (24) vanishes. The next dominate term is $\epsilon\phi_{\xi\tau}$ and it has the same order of magnitude as the term on right hand side of the same equation. We obtain

$$\mu = \epsilon^{1/3}$$

and

$$\xi' = \xi/\epsilon^{1/3} = \epsilon^{2/3}(x - c_g t).$$
$$\tau' = \tau/\epsilon^{2/3} = \epsilon^{4/3}t.$$

Equalize the order of second term with the rest of the terms in equation (24), we find that the scale for cross-stream direction remains the same, i.e.

$$\eta' = \eta = \epsilon y.$$

As a conclusion, our multiple scales looks like

$$\xi = \epsilon^{2/3}(x - c_g t), \quad \xi_1 = \epsilon^{4/3}(x - c_g t), \dots$$
$$\tau = \epsilon^{4/3}t$$

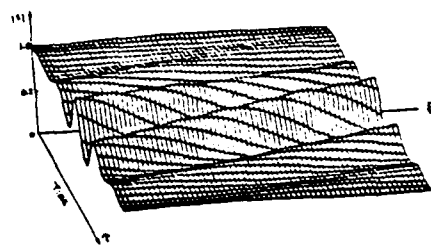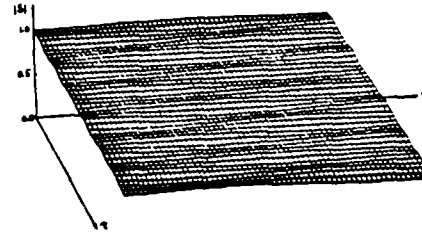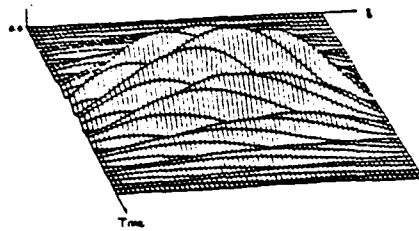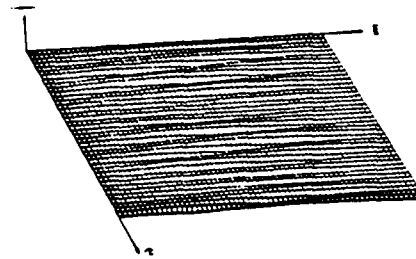**Figure 1-4** Numerical solutions for 1D equations (27 and 28), showing short and long-wave envelopes.

Based on new scales, we derived the evolution equations describing two-dimensional long-wave/short wave interaction as following:

$$iA_\tau + \lambda A_{\xi\xi} = \delta A\phi_\xi \tag{32}$$

$$\phi_{\xi\tau} + \beta\phi_{\eta\eta} = -\alpha(|A|^2)_\xi \tag{33}$$

If we substitute with β as in $1 - D$ case, i.e. define $\beta(\xi, \eta, t) = \delta\phi_\xi$, we get

$$iA_\tau + \lambda A_{\xi\xi} = BA \tag{34}$$

$$B_\tau - \beta \int_\xi^\infty B_{\eta\eta} d\xi = -\alpha(|A|^2)_\xi \tag{35}$$

A computer program is developed to solve this set of equations numerically. The results are shown in next section.

3.0 Results

(A) One-dimensional case.

Some interesting properties about the 1-D evolution equations [(27) and (28)] were
found by Koop and Redekopp. The instability is unidirectional in the sense that
the long wave cannot generate the short wave if the short wave is initially zero. For
the case of uniform wavetrain with small perturbation, the instability criterion is
independent of the long-wave amplitude. They also showed that the unstable range
is $0 < k < 2.182$ in terms of the normalized variables where k is the wave number
of small perturbation. Numerical computation was carried out for two cases (Ref.
to Figure 1 – 4), $k = 2.1$ (unstable) and $k = 2.25$ (stable) with initial conditions:

$$S = 1 + S_+ e^{i\kappa\xi} + s_- e^{-i\kappa\xi} \tag{36}$$

$$L = 0 \tag{37}$$

Near-perfect recurrence was observed in unstable case.

(B) Two-dimensional Case

The property of unidirection instability, i.e. the long wave cannot generate the
short wave if the short wave is initially zero, is preserved in 2-D case. However,
the integral term in equation (35) does make 2-D case behave differently from 1-D
case. Let us start the discussion with the long term behavior of the long waves.
From equation (33)

$$\Phi_{\xi\tau} + \beta\phi_{\eta\eta} = -\alpha(|A|^2)_\xi$$

Since the nonlinear effect is weak, we assume that ϕ behaves close to linear system,
i.e. the form of ϕ is assumed to be

$$\phi = \int_\infty^\infty \int_\infty^\infty F(k, l) \exp\left\{ i\left(k\frac{\xi}{\tau} + l\frac{\eta}{\tau} - \frac{l^2}{k} \right)\tau \right\} dk \, dl. \tag{38}$$

The stationary phase of ϕ gives relationship

$$\xi = -\eta^2/4\tau. \tag{39}$$

This means that the long waves travel leftward when they propagate outward. It
suggests that the integration in equation (35) should start from right-most bound-
ary and the integrated value can not be periodical in the stream-wise direction.

Instead of using periodical boundary conditions as we did in the 1-D case, we impose zero gradient at the boundaries. Since a uniform wavetrain with periodical perturbation can no longer be a solution, we start with a Gaussian as initial condition as shown in Figure (5). As the short wave propagates [Figures (7) and (9)], the interaction appears only in stream-wise direction. They never expand in the cross-stream direction. This is a proof of unidirection instability. On the other hand, the long waves [Figures (8) and (10)] propagate into the left half of the domain as we predicted earlier. The boundary affects can be seen in Figure (9) and (10) compared with Figures (13) and (14). They both have exactly the same initial condition except that the domain of the later case was doubled.

A couple of different initial conditions [as shown in Figure (15) and (16)] were tested. They were so designed that they are close to wavetrain but without the periodical property. The results show that the wave number of the perturbation does effect the interaction rates, i.e. smaller wave numbers have faster interaction rates as in 1-D case. But in all cases, the amplitude of the envelope of short waves and the amplitude of long waves only oscillates within a moderate range. The unstable case as in Figure (1) was not found.

4.0 Conclusion

The equations to describe the two-dimensional resonant interaction between a shallow-water gravity wave and a capillary wave have been derived. It is found that the resonant interaction occurs on a time scale $O(\epsilon^{-4/3})$ just as in the one-dimensional case. The spatial scale in stream-wise direction is also the same as in the 1-D case, namely $O(\epsilon^{-2/3})$, but the cross-stream direction is different, being $O(\epsilon^{-1})$. The dynamical characteristics of 2-D resonant interaction are quite different from the 1-D case, as illustrated by numerical examples.

5.0 References

Chu, V.T. & Mei, C.C. (1970) *J. Fluid Mech.*, 41, part 4, 873 – 887.

Djordjevic, V.D. & Redekopp, L.G. (1977) On two-dimensional packets of capillary-gravity waves. *J. Fluid Mech.*, 79:703 – 714.

Tucker, M.J. (1983) Observation of ocean waves. *Phil. Trans. R. Soc. Lond. A*, 309:371 – 380.

Whitham, G.B. (1974) *Linear and Non-Linear Waves*. Wiley: New York.

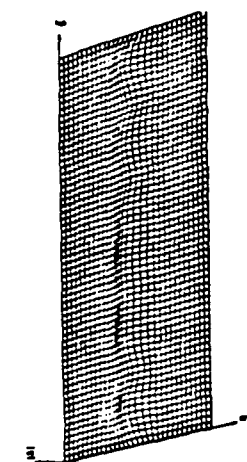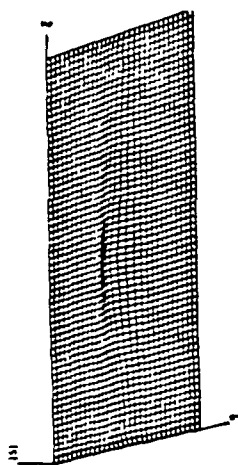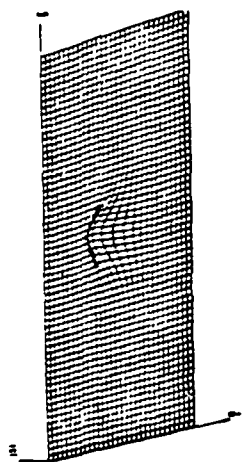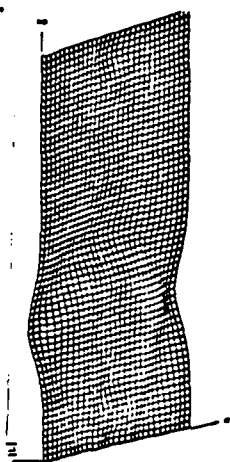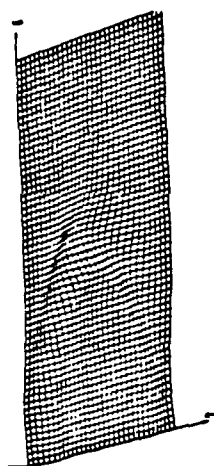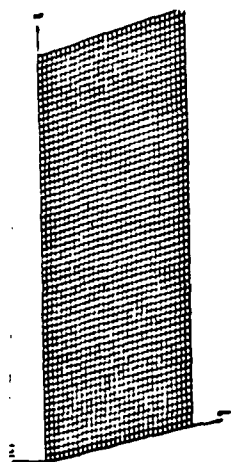Figure 5-10 Development of short-long wave interactions in the two-dimensional case of showing both short- and long-wave envelopes.
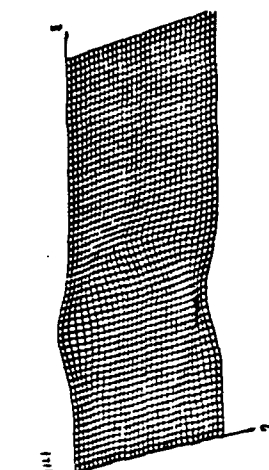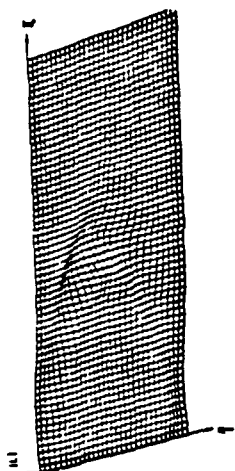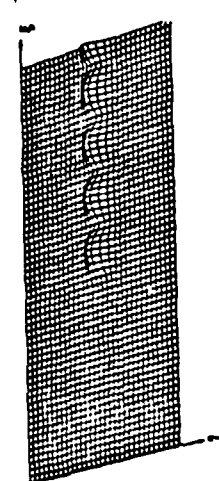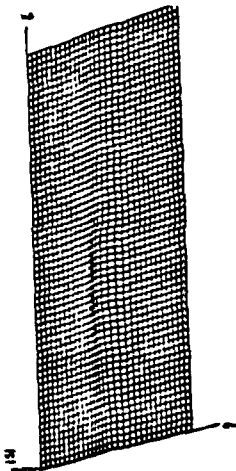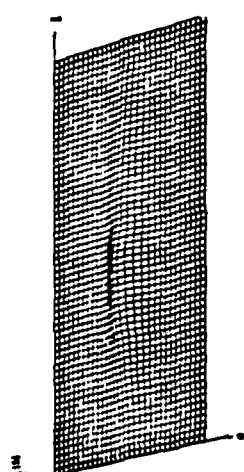
Figure 11-16 Other examples of the development of the short- and long-wave envelopes.

6.2 Optical Flow of Planar Surfaces (S. Ullman)

1. Introduction

When objects move in the environment with respect to the viewer, the images they cast upon the viewer's retina undergo complex transformations. The human visual system can interpret these transformations to recover the three-dimensional (3-D) structure of the viewed objects and their motion in space.

A substantial number of computational studies have investigated this capacity to recover structure from motion. The main two goals of these studies have been (i) to determine the conditions under which the 3-D structure can be recovered uniquely from the changing projection, and (ii) to develop methods for computing the 3-D structure and motion in space from the projected transformations.

With respect to the uniqueness problem, the main conclusion from these studies has been that for non-planar rigid objects in motion, the 3-D structure and motion are determined uniquely by the projected transformation. Results of this type have been established for both perspective and orthographic projections, and for different forms of input to the recovery process, including discrete points in discrete views, discrete points and their velocities, and a continuous velocity field.

The uniqueness results obtained in these studies depended critically upon the object being non-planar. Until recently, the problem of interpreting the motion cast by planar surfaces has remained relatively unexplored. The problem has obvious practical implications, since many surfaces are planar or nearly planar, and in these cases methods that assume non-planarity would be incorrect or unreliable. As mentioned by Waxman & Ullman (1983) and by Longuet-Higgins (1984), the analysis of motion relative to a planar surface may be relevant to certain situations of night-landing of an aircraft, where the main visual cues are roughly coplanar.

Hay (1966) was apparently the first to analyze mathematically the visual interpretation of moving planes. Hay's analysis assumed that the visual input is given in the form of two discrete views obtained from a set of points in motion.

The main result established by Hay was that the interpretation problem in this case exhibits in general a two-fold ambiguity. In addition to the moving plane that has actually cause the viewed transformation there is in general one additional "confusable" plane. The second solution is entirely different from the first in its spatial orientation and motion through space.

Similar results have been established by Tsai & Huang (1981), based on a different method of analysis. They have also shown that when the translattion

44

in space in along the normal to the surface, the ambiguity disappears and the
solution becomes unique

A recent study by Longuet-Higgins (1981) has identified an additional
condition under which the two-fold ambiguity disappears. He also provided a
different method of analysis and a different algorithm for computing the 3-D
parameters.

The current paper extends the analysis of the planar velocity field (i.e.,
the velocity field induced by a moving planar surface) in four directions. First,
it uses a different form of input to the recovery process. Instead of discrete
points, it uses continuous flow parameters such as local vorticity, shear, and
their derivatives. The use of flow parametsrs in the analysis of the optical
velocity field was suggested originally by Koenderink & van Doorn (1975)
and by Longuet-Higgins & Prazdny (1980). An analysis based entirely on this
form of input has been developed recetly by Waxman & Ullman (1983). When
this method was applied to planar surfaces, it was found empirically that
there were in general two distinct 3-D solutions. The result was not proven
mathematically, however, and the possible existence of additional solutions
was not ruled out.

An analysis based on input in the form of flow parameters is of interest for
two reasons. First, as emphasized by Koendrink & van Doorn and by Longuet-
Higgins& Prazdny, the flow parameters are convenient in the sense that they
usually have a clear geometric interpretation, and some of them are invariant
with respect to the choice of a coordinate system. Second, in analogy with
the non-planar case, it is of interest to examine the interpretation problem
for different forms of input since the analysis of each case is usually different,
leading to a different recovery method with somewhat different properties.

The second direction in which the current paper extends previous inves-
tigations is the consideration of confusable non-planar solutions. Previous
studies have assumed that the moving surface is known to be planar. and ex-
amined the number of possible planar solutions. It is of interest, however, to
determine the ambiguity of the interpretation when the moving surface is not
known in advance to be planar. It is shown here that confusable non-planar
solutions can in general be ruled out. Third, the method of analysis is differ-
ent from previous studies, leading to a different family of possible algorithms.
Finally, the information available in the orthographic rather than perspec-
tive velocity field of planar surfaces is analyzed. Since under local analysis
perspective and orthorgraphic projections become almost indistinguishable,
this analysis indicates some limits on the information that can be extracted
reliably from a local analysis of the velocity field.

2. The Velocity Field of a Moving Surface

The analysis of the planar velocity field will proceed in two steps. In the first step the projected velocity field will be expressed in terms of the 3-D shape and motion parameters of the inducing surface. The description will follow the derivation in Waxman & Ullman (1983). This step is straightforward, and the resulting 2-D velocity field is obviously uniquely determined by the 3-D parameters. The second step (Sections 3.4.5) consists of inverting this process: given the instantaneous flow field of a planar surface, the problem is to recover the unknown 3-D parameters.

2.1 Notation

Following Longuet-Higgins & Prazdny we will use a coordinate system (X, Y, Z) moving with the observer relative to the scene. The origin of the coordinate system is the vertex of the perspective projection from the scene to the image, and the Z-axis is oriented along the instantaneous line of sight. The Z-axis intercepts the viewed object at $(0, 0, Z_0)$. It is assumed that around this point the object can be described by a twice-differentiable surface (not necessarily planar) $Z(X, Y)$. Image coordinates will be denoted by lower-case letters (x, y) where $x = \frac{X}{Z}, y = \frac{Y}{Z}$. U, V, W will denote velocities in space in the X, Y, Z directions, and u, v image velocities in the x, y directions. Subscripts such as u_x, u_y will denote partial derivatives along the x and y direction.

2.2 The Instantaneous Velocity Field Around the Origin

The instantaneous motion of a rigid object can be described by six independent parameters. We will denote them by $M_i, i = 1, \ldots 6$. M_1, M_2, M_3 are the velocities along the X, Y, Z directions scaled by the distance $Z_0, M_1 = U/Z_0, M_2 = V/Z_0, M_3 = W/Z_0$. M_4, M_5, M_6 are the angular velocities around the X, Y, Z axes. The shape parameters we will use are $T_1, \ldots, T_5$, the surface orientation at the origin $T_1 = \left(\frac{\partial Z}{\partial X}\right)_0$, $T_2 = \left(\frac{\partial Z}{\partial Y}\right)_0$, and the surface curvature at the origin (scaled by Z_0) $T_3 = Z_0\left(\frac{\partial^2 Z}{\partial X^2}\right)_0$, $T_4 = Z_0\left(\frac{\partial^2 Z}{\partial Y^2}\right)_0$, $T_5 = Z_0\left(\frac{\partial^2 Z}{\partial x \partial y}\right)_0$. T_i are assumed below to be finite.

The image observables that will be used to recover the unknown 3-D shape and motion parameters are the image velocity at the origin $(0,0)$ and its first and second derivatives in the x and y directions. Rather than using these derivatives themselves, we will use linear combinations of them, as suggested originally by Koenderink & van Doorn (1975) and by Longuet-Higgins & Prazdny (1980).

We will use a set of 12 image observables denoted by $O_1,....,O_{12}$. They are defined as follows:

$$O_1 = u \qquad\qquad O_7 - u_{xx}$$
$$O_2 - v \qquad\qquad O_8 - u_{xy}$$
$$O_3 : u_x \qquad\qquad O_9 - v_{yx}$$
$$O_4 - v_y \qquad\qquad O_{10} - v_{yy}$$
$$O_5 - \tfrac{1}{2}(u_y - v_x) \qquad\qquad O_{11} \quad u'_x$$
$$O_6 = \tfrac{1}{2}(u_y - v_x) \equiv v' \qquad O_{12} - u'_y$$

The first two observables evaluated at $(0,0)$ are the image velocity at the origin in the x and y directions. The next four can be thought of as describing the deformation of a differential neighborhood around the origin. O_3 and O_4 are the rate-of-stretch along the x and y axes, O_5 measures the rate of decrease in the angle between line elements oriented along the axes, and O_6 is the local rate of rotation. $O_7 - O_{12}$ are the spatial derivatives of these variables. By evaluating explicitly the various derivatives it is straightforward to obtain the mathematical relations between the observables $O_1,...,O_{12}$ and the unknown parameters $M_1,...,M_6$ $T_1,...T_5$ (Waxman & Ullman 1983. See also a similar derivation in Longuet-Higgins & Prazdny 1980). The resulting equations are:

$$(1)$$

$$O_1 = -M_1 - M_5 \qquad\qquad O_7 = -2(M_5 + M_3 T_1) + M_1 T_3$$
$$O_2 = -M_2 + M_4 \qquad\qquad O_8 = M_4 - M_3 T_2 + M_1 T_5$$
$$O_3 = M_3 + M_1 T_1 \qquad\qquad O_9 = -M_5 - M_3 T_1 + M_2 T_5$$
$$O_4 = M_3 + M_2 T_2 \qquad\qquad O_{10} = 2(M_4 - M_3 T_2) + M_2 T_4$$
$$O_5 = \tfrac{1}{2}(M_2 T_1 + M_1 T_2) \qquad\qquad O_{11} = \tfrac{1}{2}(-M_4 + M_3 T_2 + M_2 T_3 - M_1 T_5)$$
$$O_6 = -M_6 + \tfrac{1}{2}(M_2 T_1 - M_1 T_2) \qquad O_{12} = \tfrac{1}{2}(-M_5 - M_3 T_1 - M_1 T_4 + M_2 T_5)$$

These relations can be used for the recovery of the 3-D shape and motion parameters by solving for the M_i and T_i when the O_i are given (measured in the image). Longuet-Higgins & Prazdny (1980) have shown that for non-planar patches a similar set of equations has at most three different solutions. It was found in computer simulations (Waxman & Ullman 1983) that the solution is usually unique, but an analytic uniqueness proof is still lacking. We next turn to the planar case and show that eq. (1) then have in general exactly two distinct solutions.

3. The two-fold ambiguity of planar surfaces

In the planar case the surface curvature parameters T_3, T_4, T_5 in (1)

all vanish. The resulting equations are still coupled and non-linear. and the number of distinct solutions is not immediately apparent. In this section it is shown that there are in general two distinct solutions. In addition to the surface that gave rise to the velocity field there is in general one (and only one) additional surface. engaged in a different motion. that can produce an identical velocity field.

3.1 There are in general at least two distinct solutions

When $T_3 = T_4 = T_5 = 0$ the last four equations in (1) are immediately derivable from the preceding eight and can be ignored. If the resulting system of eight equations has a solution, it also has a second solution, that is in general different from the first. This claim is established by giving explicitly a second solution in terms of the first. Let $(M_1, ..., M_6, T_1, T_2)$ be a solution to (1) (with $T_3 = T_4 = T_5 = 0$). A second solution $(\overline{M}_1, ...\overline{M}_6, \overline{T}_1, \overline{T}_2)$ can be derived explicitly as follows:

$$
\begin{aligned}
\overline{T}_1 &= -M_1/M_3 \\
\overline{T}_2 &= -M_2/M_3 \\
\overline{M}_1 &= -T_1 M_3 \\
\overline{M}_2 &= -T_2 M_3 \\
\overline{M}_3 &= M_3 \\
\overline{M}_4 &= M_4 - M_2 - M_3 T_2 \\
\overline{M}_5 &= M_5 + M_1 + M_3 T_1 \\
\overline{M}_6 &= M_6 + M_1 T_2 - M_2 T_1
\end{aligned}
\tag{2}
$$

This solution exists provided that $M_3 \neq 0$. Ths case $M_3 = 0$ is examined in section 4 below. The two solutions are dual in the sense that either one can be used in (2) to obtain the other. In the two solutions the directions of the translation vector and the surface normal are interchanged. If v, n are the translation vector and surface normal in the first solution, then the second $\overline{v}$ points in the direction of n, and $\overline{n}$ in the direction of v.

We next turn to show that there are no more than two distinct solutions. This will be done in two stages. Section 3.2, which is the main step in the proof, shows that if two planar surfaces have identical velocity fields then they have the same value of M_3 (velocity along the line of sight). Section 3.3 establishes that there are at most two solutions that share the same value of M_3.

3.2 Two planes that induce the same velocity field have the same value of M_3

Let π and $\bar\pi$ be two planes engaged in motions $(M_1, ..., M_6)$ and $(\bar M_1, ..., \bar M_6)$ respectively, that induce identical velocity fields (i.e., identical observables in (1)). We can assume that the two planes intersect. Otherwise, $T_1 = T_1, T_2 = T_2$, which implies either (i)$(M_1, ..., M_6) = (M_1, ..., \bar M_6)$, or (ii) all of the translation components $M_1, M_2, M_3, M_1, M_2, M_3$, vanish. This latter case, corresponding to pure rotation, is uninteresting since no 3-D information is conveyed by the changing image. Let ℓ be the intersection of π and $\bar\pi$. This special line in space participates in two different motions and induces the same (linear) velocity fields. Without loss of generality we can re-orient the coordinate system so that the projection of ℓ coincides with the x axis. The line ℓ can be described now by $Z = ax + Z_0$. The main step in the proof is to consider the intersection line ℓ instead of the two planes. This line has the property that when it participates in motion $(M_1, ..., M_6)$ or $(\bar M_1, ..., \bar M_6)$, it induces identical velocity fields.

Consider the situation in which the planes no longer exist, only the single line ℓ is moving in space. Let it move with the 3-D motion parameter $(M_1 - \bar M_1, ..., M_6 - \bar M_6)$. From the original coincidence between the velocity fields of π and $\bar\pi$, and since ℓ lies on both planes, it follows that the velocity field projected by ℓ now vanishes, i.e., it satisfies at the origin the equations:

$$v = 0 \qquad\qquad u = 0$$

$$\frac{\partial v}{\partial x} = 0 \qquad\qquad \frac{\partial^2 v}{\partial x^2} = 0 \qquad\qquad (3)$$

$$\frac{\partial u}{\partial x} = 0$$

We have transformed the problem of the moving planes into a problem concerning a moving straight line. The question is: Under what conditions the velocity field of a moving line, as expressed in eq. (3), vanishes? In the re-oriented coordinate system let us denote

$$M_1 - \bar M_1 = V_x \quad M_2 - \bar M_2 = V_y \quad M_3 - \bar M_3 = V_z$$
$$M_4 - \bar M_4 = w_x \quad M_5 - \bar M_5 = w_y \quad M_6 - \bar M_6 = w_z$$

The five equations in (3) can be expressed in terms of the six motion parameters $(V_x, V_y, V_z, w_y, w_z)$. The derivation is somewhat lengthy, but straightforward. The resulting equations are:

$$V_x + Z_0 w_y = 0$$
$$V_y - Z_0 w_x = 0 \qquad\qquad (4)$$
$$V_z + a V_x = 0$$
$$a V_x + Z_0 w_y = 0$$
$$a V_y - Z_0 w_z = 0$$

From which it follows that:

$$V_j \quad 0 \; V_z \quad 0 \; w_u \quad 0 \; V_y \quad Z_{0u}, \quad \alpha w_j \quad \alpha_z$$

In terms of the original planes π, π, the implication of $V_z - 0$ is that $M_x = \overline{M}_x$. In the particular coordinates system in which ℓ projects onto the x-axis, it is also true that $M_1 \quad \overline{M}_1$ and $M_5 - \overline{M}_5$.

3.3 The number of distinct solutions cannot exceed two

Since all the possible planar solutions share the same value of M_3, the proof will be completed by showing that for a given M_3 there are at most two distinct solutions. We proceed along the following plan. If π is a planar solution let ℓ be now the intersection line of π with the frontal plane $Z = Z_0$. We will call ℓ the "tilt line" of π. If we re-orient the coordinate system so that ℓ runs along the x-axis, then in the new coordinate system $T_1 = 0$. From eq. (1), in this reoriented coordinate system $O_3 = M_3$. It can be observed from eq.(1) that, if we exclude solutions in which $T_1 = 0$ and $M_1 = 0$ simultaneously, the a fixed M_3 and $T_1 = 0$ determine the solution uniquely. We will show that there are at most two orientations of the coordinate system for which $O_3 = M_3$. This will imply that (except for two special cases that will be examined separately) there are at most two distinct tile lines and therefore at most two distinct solutions. We will assume here $M_3 \neq 0$, the case $M_3 = 0$ is examined in section 4. It is convenient for the proof to assume that the velocity field satisfies initially $O_3 = O_4$. This can be assumed without loss of generality, since it is always possible to satisfy the assumption by re-orienting the coordinate system (choosing new x, y coordinates) in the following manner. Let us rotate the coordinate system along the line of sight by an angle β and denote the eight observables in the new coordiante system by $(\overline{O}_1, ..., \overline{O}_8)$. We will determine an angle β such that following the rotation $\overline{O}_3 = \overline{O}_4$. In the rotated coordinate system:

$$\overline{O}_3 = O_3 cos^2\beta + O_4 sin^2\beta + 2O_5 sin\beta cos\beta$$
$$\overline{O}_4 = O_4 cos^2\beta + O_3 sin^2\beta - 2O_5 sin\beta cos\beta \qquad (5)$$

to obtain $\overline{O}_3 = \overline{O}_4$, β must satisfy

$$(O_3 - O_4)cos2\beta + 2O_5 sin2\beta = 0$$

assuming $O_5 \neq O$, β is determined by

$$tan2\beta = \frac{O_4 - O_3}{2O_5} \qquad (6)$$

There will always be a solution for β (Not necessarily unique. The case $O_2 = 0$ will be examined separately.) We can assume therefore that we have initially a coordinate system in which $O_2 = O_1$. We now rotate the coordinate system by a new angle α, and denote again the observables before the rotation be $(O_1,...,O_n)$ and following the rotation by $(O_1,...,O_n)$. We seek an angle α for which $O_3 = M_3$. In general $O_3 = O_3 \cos^2\alpha + O_4 \sin^2\alpha - 2O_5 \sin\alpha\cos\alpha$. But since $O_2 = O_4$

$$O_3 = O_3 - 2O_5 \sin\alpha\cos\alpha \qquad (7)$$

assuming $O_5 \neq 0$

$$sin2\alpha = \frac{M_3 - O_3}{O_5}$$

There are four solutions for α in the range $[0, 2\pi]$, α_1, α_2, $\alpha_1 + \pi$, $\alpha_2 + \pi$. The solution $\alpha_1, \alpha_1 + \pi$ are equivalent, they give the same solution for $(M_1, ..., M_6, T_1, T_2)$.

This establishes the claim for the general case. Two special cases that were excluded from the proof can be analyzed in a similar manner. In the first case there is a solution for which $T_1 = 0$ and $M_1 = 0$ simultaneously. In this case the velocity field will have a single direction along which $O_3 = M_3$. M_1, M_3, M_5, T_1 are then determined uniquely, but there are two solutions for M_2, M_4, T_2, in (1). In the second case $O_5 = 0$ for every orientation of the coordinate system. In this case there are two solutions, in one $T_1 = T_2 = 0$ (frontal plane) and in the other $M_1 = M_2 = 0$ (motion along the line of sight). In all of these cases the velocity field admits at most two planar interpretations.

4. Unique Solutions

In the previous section it was shown that the velocity field of a moving plane is compatible, in general, with exactly one additional moving plane.

There are two special cases under which the two-fold ambiguity disappears, and a degenerate case under which the surface orientation cannot be recovered. The discussion of these cases will be brief, since one of these cases is discussed in Ullman & Waxman (1983) and the other in Longuet-Higgins (1984).

The degenerate case arises when $M_1 = M_2 = M_3 = 0$. The motion in this case is pure rotation. The motion parameters can be recovered, but the surface orientation remains ambiguous.

One unambiguous case arises when there is no relative velocity along the observer's line of sight, i.e. $M_3 = 0$ (but M_1 and M_2 are not both zero). In this

case the motion equations can be solved explicitly and the solution is unique
The second unambiguous case arises when $M_5 \neq 0$ and the observer moves
directly towards or away from the surface. If the original motion satisfies
$M_1 M_3 = T_1, M_2 M_3 = T_2$ then it can be seen from equation (2) that the
second solution coincides with the first. The latter condition is discussed in
Longuet-Higgins. (1984) He also suggested an additional condition that can
be used to resolve the ambiguity, by using an extended region of the plane
rather than the local velocity field.

In the previous section we have left out the case $O_5 = 0$. Inspection
of the equations reveal that this case falls into one of the categories already
discussed. One case where $O_5 = 0$ is obtained when $M_1 = M_2 = M_3 = 0$,
the ambiguous case discussed above. A second case is when $M_1 = M_2 = 0$
(but $M_3 \neq 0$) and $T_1 = T_2 = 0$. In this case the motion is along the surface
normal, and the solution is unique. In all other cases there are two distinct
solutions.

4.1 Summary

The ambiguous and non-ambiguous solution can be summarized as fol-
lows.

1. In the pure rotation case the motion can be recovered but the surface
   orientation remains ambiguous.
2. If there is no relative velocity along the line of sight ($M_3 = 0$ but $M_1 M_2 \neq$
   0), or if the motion satisfied $M_1/M_3 = -T_1$, $M_2/M_2 = -T_2$ (motion
   perpendicular to the plane), then the solution is unique.
3. In all other cases the local velocity field has a two-fold ambiguity. The
   second solution can be derived in terms of the first by equation (2).

5. The exclusion of non-planar solutions

We have seen in the previous section that the velocity field of a moving
plane has in general one additional planar interpretation. That is, if the
moving surface is known to be a plane, there are in general two distinct
solutions. The possibility remains, however, that when nothing is known in
advance about the surface, there are additional non-planar interpretations.
In this section the question of non-planar ambiguities is considered. In other
words, the question is whether in addition to the two planar interpretations,
non-planar solutions are also possible.

Let π be a moving plane and let μ denote the vector $(M_1, ..., M_6)$ of its
motion parameters. Suppose that the twice-differentiable surface P is non-
planar around the origin and that it induces the same velocity field as π, and

let p denote the motion parameters of P. It is shown that if the velocity fields of P and π coincide in a neighborhood around the origin, then P is in fact planar at the origin. That is, if T_3, T_4, T_5 are the surface curvature parameters of P at the origin then $T_3 = T_4 = T_5 = 0$.

Without loss of generality we can assume that both surfaces pass through the point $(0,0, Z_o)$. We can then distinguish between two cases.

Case 1: π is tangent to P at $(0,0, Z_0)$. In this case π and P have the same values for T_1, T_2. From the original equation (1) it can be verified that the only ambiguous configuration in this case is when M_1 and M_2 in (1) both vanish. That is, the motion is directed along the line of sight. In all other cases P must coincide at the origin with π, and satisfy $\overline{T}_3 = \overline{T}_4 = \overline{T}_5 = 0$.

Case 2: π intersects P along some space-curve c. Let γ be the projection of c on the image plane. Assume first that near the origin γ is not a straight line segments. Longuet-Higgins (1984) has shown that given the image velocities of four coplanar points (no three of which are colinear in the image), the 3-D motion parameter can be recovered up to the two-fold ambiguity discussed in the previous section.

The implication is that the motion p coincides either with μ or with $\overline{\mu}$, the motion of the dual solution to μ. In either case we obtain that planar and non-planar surfaces with identical motion parameters produce an identical velocity field. From the original equation (1) it can be verified that the only ambiguous configuration in this case arises again when $M_1 = M_2 = 0$. If M_3 also vanishes, (the pure rotation case), the situation is inherently ambiguous, as discussed in the previous section. If $M_3 \neq 0$ then π is in fact tangent to P, as in the previous case.

The only remaining possibility is that near the origin γ is a straight line segment. In this case we can use the results of section 3.2. Without loss of generality γ can be assumed to lie along the x axis. Let M_i, T_i denote the motion and shape parameters of π, $\overline{M}_i, \overline{T}_i$ of P. From 3.2, it follows that $M_3 = \overline{M}_3, M_1 = \overline{M}_1, M_5 = \overline{M}_5$. In addition $T_1 = \overline{T}_1$ and $T_3 = \overline{T}_3$ since both are measured along c which is straight line in space common to π and P, and $\overline{T}_1 = 0$ since π is planar. $\overline{T}_4, \overline{T}_5$ can now be analyzed using eq. (1). The equality of the observable O_9 in (1) implies $M_2 T_5 = \overline{M}_2 \overline{T}_5$, and since $T_5 = 0$, $\overline{M}_2 \overline{T}_5 = 0$. Similarly O_9 and O_{12} together imply $\overline{M}_1 \overline{T}_4 = 0$, and also $M_1 \overline{T}_4 = 0$ since $\overline{M}_1 = M_1$. If $M_1 \neq 0$, $\overline{T}_4 = 0$. From O_{10} together with O_8 it now follows that $M_1 T_5 = M_1 \overline{T}_5$ and therefore $\overline{T}_5 = 0$. The case $M_1 = 0$ can be analyzed in a similar manner.

The final conclusion is that $\overline{T}_3 = \overline{T}_4 = \overline{T}_5 = 0$ except for the case $M_1 = M_2 = 0$ (motion along the line of sight).

In conclusion, we can distinguish between two cases. If the relative motion happens to be along the line of sight, then the local velocity field of a plane π is also compatible with any non-planar surface with the same relative motion parameters, and whose tangent plane coincides with π. Unlike the planar two-fold ambiguity, this ambiguity is local, and can be resolved by inspecting a larger region of the plane.

In the more general case, in which the motion component parallel to the image plane does not vanish, non-planar solutions can be ruled out, and the only remaining ambiguity is the two-fold planar ambiguity.

6. Algorithm

The proof in section 3 although not entirely constructive, leads to a possible algorithm for computing the planar solutions. The method will be outlined briefly. A different algorithm has been suggested by Longuet-Higgins (1984).

We begin by rotating the coordinate system by an angle θ to obtain $\overline{O}_3 = \overline{O}_4$ (where $\overline{O}_i$ are the observables in the rotated coordinated system). From (6) we obtain

$$tan2\theta = \frac{O_4 - O_3}{2O_5}$$

(Provided that $O_5 \neq 0$). We can therefore obtain a solution (non unique) for $sin\theta, cos\theta$. It is a straightforward computation to then compute the observables $(\overline{O}_1, ..., \overline{O}_8)$ (Waxman & Ullman, 1983).

The planar motion equations (first eight equations in (1)) can be viewed as eight linear equations in 12 unknown: $M_1, ..., M_6$, and $X_1, ..., X_6$, when $X_1 = M_1 T_1$, $X_2 = M_2 T_2$, $X_3 = M_1 T_2$. $X_4 = M_2 T_1$, $X_5 = M_3 T_1$, $X_6 = M_3 T_2$. The eight equations are linearly independent. Furthermore, they can be divided into four groups of two equations.

(M_3, X_1, X_2) appear only in equations (3,4).

(M_6, X_3, X_4) appear only in (5,6).

(M_1, M_5, X_5) appear only in (1,7).

(M_2, M_4, X_6) in (2,8).

As a result, each set of unknowns can be solved up to a single scalar:

$$(M_3, X_1, X_2) = (m_3, x_1, x_2) + \alpha(\overline{m}_3, \overline{x}_1, \overline{x}_2)$$
$$(M_6, X_3, X_4) = (m_6, x_3, x_4) + \beta(\overline{m}_6, \overline{x}_3, \overline{x}_4) \quad (8)$$
$$(M_1, M_5, X_5) = (m_1, m_5, x_5) + \gamma(\overline{m}_1, \overline{m}_5, \overline{x}_5)$$
$$(M_2, M_4, X_6) = (m_2, m_4, x_6) + \delta(\overline{m}_2, \overline{m}_4, \overline{x}_6)$$

the m, m, x, x, $(i = 1, ..., 6)$ are determined by solving sets of two equations in three unknowns. The scalars α, β, γ, δ remain to be determined. Since $O_3 = O_4$, it follows from (1) that $M_1 T_1 = M_2 T_2$, i.e. $X_1 = X_2$, therefore $x_1 - \alpha x_1 = x_2 - \alpha x_2$ and hence α is determined. Two possible values for β, γ, δ are obtained from the equations

$$X_1 X_2 = X_3 X_4$$
$$M_3 X_1 = M_1 X_5 \tag{9}$$
$$M_3 X_2 = M_2 X_6$$

respectively. Finally we choose a value for $(\alpha, \beta, \gamma, \delta)$ to satisfy the fourth independent relation $M_2 X_1 = M_1 X_4$. There will be at least one such set of values for $(\alpha, \beta, \gamma, \delta)$, and at most two. When one solution is known, the second can be found immediately using equations (2).

7. The orthographic velocity field

Previous sections have established that the parameters of a moving plane can be determined up to the two-fold ambiguity from an arbitrarily small patch near the origin. When the viewed surface patch is small, perspective effects become small, and the recovery process may become unreliable. It therefore becomes of interest to analyze the case of orthographic projection where perspective effects play no role. (In orthographic projection a space point X, Y, Z, projects to an image point $x = X, y = Y$.)

7.1 The orthographic velocity field of non-planar surfaces

Let S_1 be a non-planar surface moving in space. For the orthographic case it can be assumed that S_1 is fixed at the origin (O, O, Z_0) since the translation components are immediately recoverable. The rotation of S_1 can always be decomposed into the sum of two components: a rotation with angular velocity w (assumed to be non-zero) about an axis lying somewhere in the frontal plane (x-rotation) and a component (z-rotation) about the line of sight Z with angular velocity θ.

A second surface S_2 is said to be a *depth scaling* of S_1 if:

1. For every point (x, y, z) on S_1, (x, y, kz) is a point on S_2 for some constant k $(k \neq 0)$.
2. The rotations w_1 and w_2 are around the same axis, and $w_2 = w_1/k$.
3. $\theta_1 = \theta_2$

For non-planar surfaces the following proposition can be established. If S_1 is a possible rigid interpretation of a given orthographic velocity field, then

S_2 is another possible interpretation if and only if it is obtained from S_1 via depth scaling.

Note that if $\theta_1 \quad \theta_2 = 0$, then S_1 and S_2 have different instantaneous axes of rotation in space. The orthographic velocity field therefore does not determine uniquely the rotation axis in space.

Since our concern here is primarily with planar surfaces, the proof will be omitted. it can be found in (Ullman 1983).

7.2 The orthographic velocity field of planar surfaces

In the planar case the twofold ambiguity of planar surfaces is combined with the inherent depth-scaling ambiguity of orthographic projection. As a result the orthographic projection of a planar surface admits two interpretations, each defined up to depth scaling.

For simplicity of the analysis we can assume that in the planar velocity field all the velocity vectors are parallel to the x axis. (If the inducing object's Z-rotation is θ, then by rotating the observed velocity field by $-\theta$ all the velocity vectors will become parallel. Their direction can be taken as the x-axis.)

The velocity field $u(x, y)$ $v(x, y)$ now has the form:

$$u(x, y) = \alpha x + \beta y$$
$$v(x, y) = 0 \tag{10}$$

If (w_x, w_y, w_z) is the angular velocity vector of the rotating surface (assumed to be non-zero) then:

$$w_y z - w_z y = \alpha x + \beta y$$
$$w_z x - w_x z = 0 \tag{11}$$

One solution to these equations arises when $w_z = 0$. This implies $w_x = 0$ (if z is not identically zero), and $z = \frac{1}{w_y}(\alpha x + \beta y)$. This solution corresponds to a plane rotating about the vertical axis.

If $w_x \neq 0$ then $w_z \neq 0$ also and $z = \frac{w_x}{w_z} x$. This solution is also a plane, with a tilt line along the x axis.

These two possible interpretations cannot be resolved on the basis of the instantaneous velocity field. How much additional information is required to guarantee a unique solution? For non-planar objects, it can be shown (Ullman 1983) that one additional view is sufficient to remove the depth-scaling ambiguity. For planar objects, the problem is open.

The orthographic velocity field is thus inherently more ambiguous than the perspective one. Instead of two solutions there are two families of solutions, each determined up to depth scaling.

The additional ambiguity of the orthographic velocity field implies that under local analysis (i.e. using a small surface patch) the 3-D recovery process is not entirely stable. Aspects of the 3-D structure that are invariant under depth scaling are expected to be more stable than others. For planar surfaces these invariants include the orientation of the tilt line and the rotation component around the line of sight. Parameters that are not invariant under depth scaling such as surface slant are expected to be less robust.

8. Summary

1. The velocity field of a planar surface exhibits in general a two-fold ambiguity. In addition to the moving plane that has actually induced the viewed transformation there is one additional, and in general entirely different, planar solution.

2. There are special cases in which the interpretation of the local velocity field becomes unique. These cases are (i) $M_3 = 0$ (but $M_1 M_2 \neq 0$), and (ii) motion directed towards or away from the surface. A degenerate case arises for pure rotation. In this case the motion parameters can be determined, but the 3-D structure remains undetermined.

3. Additional non-planar solutions are in general excluded. The exception is the case of motion directed parallel to the surface normal.

4. The two planar solutions can be computed from the eight kinematic observables, using the algorithm in section 6.

5. If one of the two planar solutions is known, the dual solution can be expressed in terms of the first using eq. 2.

6. In the orthographic case there are, instead of two solutions, two families of solutions, each determined up to depth-scaling. It is expected that for the perspective case only 3-D parameters that are invariant under depth-scaling would be robust under local analysis.

The results explored in this paper are theoretical in nature. They set some limits on the performance of any motion perceiving device. It is unknown, however, to what degree the human visual system can approach these theoretical limits. It may be of interest, therefore, to test psychophysically some of the implications of the above analysis. For example:

—Can subjects interpret the velocity field of planar surfaces? (i.e., make some reliable judgements of relative motion parameters and surface orienta-

tion). Results in (Gibson et. al. 1959) indicate that under some condition this is possible, although the accuracy is probably not very high

Can they interpret, at least to some degree, the planar velocity fields under brief presentation? If they do, how do they handle the inherent twofold ambiguity?

--Can observers interpret orthographic velocity fields? Can they recover, for example, the tilt of one or both planar solutions?

Answers to these questions may give us a better insight into the processing employed by the human visual system in the recovery of structure from motion.

Acknowledgements

References

Braunstein, M.L. 1976. *Depth Perception Through Motion* New York: Academic Press.

Clocksin, W.F. 1980. *Perception of surface slant and edge labels from optical flow: a computational approach. Perception*, 9(3), 253-269.

Gibson, E.J., Gibson, J.J., Smith, O.W. & Flock.H. 1959. Motion parallax as a determinant of perceived depth. *J. Exp. Psychol.*,8(1), 40-51.

Gibson, J.J. 1950. *Perception of the Visual World*. Boston: Houghton Mifflin.

Green, B.F. 1961. Figure coherence in the kinetic depth effect. *J.Exp. Psychol.*, 62(3), 272-282.

Hay, C.J. 1966. Optical motions and space perception - an extension of Gibson's analysis. *Psychological Review*, 73, 550-565.

Helmholtz, H. von (1925) *Treatise on Physiological Optics* (ed. J.P.C. Southall) New York: Dover.

Hoffman, D.D. & Flinchbaugh, B.E. 1982. The interpretation of biological motion. *Biol. Cybern.*, 42, 195-204.

Johansson, G. 1964. Perception of motion and changing form. *Scand. J. Psychol.*, 5, 181-208.

Koenderink, J.J. & van Doorn, A.J. 1976. Local structure of movement parallax of the plane. *J.Opt. Soc. Am.*, 66, 717-723.

Koenderink, J.J. & van Doorn, A.J. 1975. Invariant properties of the motion parallax field due to the motion of rigid bodies relative to the observer. *Optica Acta*, 22 (9), 773-791.

Lee, D.N. 1980. The optic flow field: the foundation of vision. *Phil. Trans. Roy. Soc. Lond., B.* **290** 169-179

Longuet-Higgins, H.C. 1982. The role of the vertical dimension in stereoscopic vision. *Perception.* **11**, 377-386

Longuet-Higgins, H.C. 1984. The visual ambiguity of a moving plane. *Proc. Roy. Soc. Lond., B.* **223**. 165-175.

Longuet-Higgins, H.C. & Prazdny, K. 1980. The interpretation of a moving retinal image. *Proc. Roy. Soc. Lond., B,.* **208**, 385-397.

Miles, W.R. 1931. Movement interpretations of the silhouette of a revolving fan. *American Journal of Psychology* **43**, 392-405.

Prazdny, K. 1980. Egomotion and relative depth map from optical flow. *Biol. Cyber.,* **36**. 87-102.

Tsai, R.Y. & Huang, T.S. 1982. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. University of Illinois at Urbana-Chamgaign Coordinated Science Laboratory Report R-921.

Tsai, R.Y. & Huang, T.S. 1981. Estimating 3-D motion parameters of a rigid planar patch. Research Report R-922, Coordinated Science Laboratory, University of Illinois.

Ullman, S. 1983. Recent computational studies in the interpretation of structure from motion. In:*Human and Machine Vision.* A. Rosenfeld & J. Beck (ed.), New York: Academic Press.

Ullman, S. 1979. *The Interpretation of Visual Motion.* Cambridge & London: MIT Press.

Ullman, S. 1979a. Relaxation and constrained optimization by local processes. *Comp. Graph. Im. Proc.,* **9(6)**, 115-125.

Wallach, H., O'Connell, D.N. 1953. The kinetic depth effect. *Journal of Experimental Psychology,* **45**, 205-217.

Waxman, A.M. & Ullman, S. 1983. Surface structure and 3-D motion from image flow: a kinematic analysis. Univ. of Maryland Computer Science Technical Report CS-TR-1332.

Webb, J.A. & Aggarwal, J.K. Visually interpreting the motions of objects in space. *Computer,* **14(8)**, 40-49.

6.3 Rigidity and Smoothness of Motion (S. Ullman and A. Yuille)

The smoothness assumption in measuring visual motion

The problem of measuring visual motion is, in many situations, undercon-strained. That is, the information in the changing image is insufficent to determine the motion uniquely. This indeterminateness is often referred to as the "aperture problem" (for example Marr and Ullman, 1982). An additional constraint is therefore required for resolving the ambiguity and determining the velocity field uniquely. An important constraint that has been proposed for solving the problem is a smoothness constraint. (Fennema and Thomp-son 1979, Horn and Schunck 1981, Hildreth 1984, Nakayama and Silverman 1986). When applied to the problem of measuring the velocity of image con-tours, this smoothness assumption has been formulated by Hildreth in the following manner. Of all the velocity fields that are consistent with the trans-forming image, select the velocity field that minimizes the overall variation. That is, if v denotes the velocity at a point, and s the arc length along the curve, the preferred velocity field is the one that minimizes the integral:

$$\int \|\frac{dv}{ds}\|^2 \, ds$$

along the curve (the principle of least variation). It has been shown that this method of determining the velocity field gives good results under a wide

range of conditions, and it seems to correspond in many cases to the velocity

field perceived by human observers (Hildreth 1986, Nakayama and Silverman

1986).

Justifying the smoothness assumption.

The main rationale raised in support of the smoothness assumption is that

the velocity field induced by smooth contours in motion is expected to be

smooth. This argument is insufficent, however, for justifying the principle of

least variation. The smoothness assumption implies that if the object is known

to be smooth, the velocity field should not contain discontinuities. It does not

imply, however, that "the smoother the better". Consider, for instance, two

points in the image separated by a small distance, and moving with image

velocities V_1 and V_2. One may argue that if they lie on the same surface,

extremely high values of $\|V_1 - V_2\|$ are unlikely, because of some physical

limitations on the motion of objects in space. But why should one assume

that a relative velocity of, say, 0.2 deg/sec is less likely than 0.1 deg/sec? Such

a preference is assumed by the least variation principle (that favors velocity

fields in which the difference in velocity between neighboring points is as small

as possible), but cannot be defended only on the basis of a general smoothness

argument. Clearly, a stronger constraint than just the general smoothness of

surfaces is required.

Another general property that seems to provide a useful constraint in the analysis of visual motion is rigidity. Computational studies have shown that the 3D stucture of rigid and quasi-rigid objects can be recovered by looking for the most rigid interpretation possible of the changing image (Ullman 1979, 1984).

In this paper we argue that local rigidity of the object and the principle of least variation in the velocity field are related. To investigate this we consider a rod moving in three-dimensional space. The rod is allowed to move in a semi-rigid manner: it can rotate and translate freely and is also allowed a certain expansion. We calculate the relative velocity between the two endpoints of the rod, as projected on the image plane. The calculations show that under a wide range of conditions this velocity distribution is peaked at zero velocity and decreases monotonically at higher velocities. As a result, the projected velocity field of points linked rigidly together is likely to be consistent with the principle of least variation in the velocity field. There are two factors contributing to this strong bias towards small differential velocities in the projected velocity field: (i) the rigidity of the link between the two points and (ii) the effect of projection from 3-D to the 2-D image plane.

If we imagine an object being made up of a collection of semi-rigid rods

this analysis suggests that the projected velocity field is likely to be smooth. We conclude that the principle of least variation is a reasonable constraint that can be justified for the projected velocity field of rigid or locally rigid objects.

Section 1. The two-dimensional case.

First we consider a rigid rod moving in two dimensions and being projected onto a line. Because it is rigid the rod's velocity can be split into rotational and translational components. For our purposes the translational component can be ignored as it will not contribute to the derivative of the velocity field on the line.

Let the projected length of the rod be R and the real length be r, the angle to the vertical be θ and the angular velocity be ω. Then the projected velocity distribution $\Phi(u)$ is

$$\Phi(u) = \int \delta(R - r\sin\theta)\delta(u - r\omega\cos\theta)\Phi_r(r)\Phi_\omega(\omega)drd\omega d\theta \qquad (1.1)$$

where δ denotes the Dirac delta function, $\Phi_r(r)$ and $\Phi_\omega(\omega)$ are the distribution of r and ω respectively. $r\sin\theta$ and $r\omega\cos\theta$ are the projected lengths and velocity respectively. We assume that rods are equally likely to have any length between 0 and r_{max} and set $\Phi_r(r) = k$ between 0 and r_{max} (with $kr_{max} = 1$).

We can integrate with respect to θ using the first delta function. This gives (see Appendix 1)

$$\Phi(u) = \int_{r=0}^{r=r_{max}} \frac{1}{\sqrt{r^2 - R^2}} \delta(u - \omega\sqrt{r^2 - R^2})\Phi_\omega(\omega)dr\,d\omega. \qquad (1.2)$$

We integrate with respect to r to obtain

$$\Phi(u) = \int_{\omega=\omega_{min}}^{\infty} \frac{\Phi_\omega(\omega)d\omega}{(R^2\omega^2 + u^2)^{\frac{3}{2}}} \qquad (1.3)$$

where $\omega_{min} = u/\sqrt{r_{max}^2 - R^2}$. This lower bound comes from the delta function integration.

Now we examine the behavior of $\Phi(u)$. If we differentiate it with respect to u we get

$$\frac{\partial\Phi}{\partial u} = -\int_{\omega=\omega_{min}}^{\infty} \frac{u\Phi_\omega(\omega)}{(R^2\omega^2 + u^2)^{\frac{5}{2}}}d\omega - \frac{\Phi_\omega(\omega_{min})}{(r_{max}^2 - R^2)^2(R^2\omega_{min}^2 + u^2)}. \qquad (1.4)$$

So, *for any probability distribution* $\Phi_\omega(\omega)$ (which by definition must be a positive function) the two terms on the right hand side of (1.4) must be negative. $\Phi(u)$ must decrease strictly monotonically with u and has a unique maximum at $u = 0$.

Thus whatever the distribution of the rotation the most likely projected velocity is zero. This result is similar to that obtained by Ullman (1979) for the motion of dots. It was shown that if the motion of a random dot is

described by an isotropic probability distribution function in three dimensions

then the probability distribution for the projected two-dimensional motion is

peaked at 0.

Note that our result is independent of whether we choose an upper bound

for r. This can easily be seen by setting $r_{max} \mapsto \infty$ and $\omega \mapsto 0$ in (1.4). In

fact having an upper bound for r makes $\partial\Phi/\partial u$ more negative and makes

$\Phi(u)$ decay faster. Henceforth we assume for simplicity that there is no upper

bound for r.

To see the connection with the smoothness assumption observe that in

the limit Hildreth's smoothness measure can be expressed as

$$\int \|\frac{dv}{ds}\|^2 \, ds = \sum (\frac{u}{R})^2 \tag{1.5}$$

where the sum can be taken over a set of rigid rotating rods. The results above

show that for each rod considered independently the probability distribution

of the u is peaked at zero. Thus, with this independence assumption, we

argue that the most likely distribution of the whole contour is the one that

minimizes (1.5). We will relax the independence assumption in section 4.

Section 2. The three-dimensional case.

We now extend the analysis to a rod moving in 3-space. Consider a rod

of length r projected into the image plane, which has unit normal vector $\vec{k}$.

The rod's direction in 3-D space is denoted by $\vec{r}$ which is a unit vector in the direction $\hat{r}$. The rod is projected into a vector $\vec{r}_p$ where

$$\vec{r}_p = \vec{r} - (\vec{r} \cdot \vec{k})\vec{k}. \tag{2.1}$$

The rod's rotation is described by a vector $\vec{\omega}$, where ω is the frequency and $\hat{\omega}$ is the axis of rotation. The velocity $\vec{v}$ is given by

$$\vec{v} = \vec{\omega} \times \vec{r}. \tag{2.2}$$

The projected velocity field $\vec{v}_p$ is given by

$$\vec{v}_p = \vec{v} - (\vec{v} \cdot \vec{k})\vec{k}. \tag{2.3}$$

We consider rods with projected length R and projected velocity $\vec{U}$. We are interested in the projected velocity distribution $\Phi(\vec{U})$. Since the projected velocity distribution is rotationally symmetric we can express this as $U\Phi(U)$ (see Appendix 1). If the rods' length distribution is $\Phi_r(\vec{r})$ and their rotation distribution is $\Phi_\omega(\vec{\omega})$ then the projected velocity distribution is given by

$$\Phi(u) = \int \delta(R - |\vec{r}_p|)\delta(U - |\vec{v}_p|)\Phi_r(\vec{r})\Phi_\omega(\vec{\omega})d\vec{r}d\vec{\omega}. \tag{2.4}$$

Care is needed in specifying the domain of integration of (2.4). From (2.2) we see that only the component of $\vec{\omega}$ perpendicular to $\vec{r}$ contributes to the

velocity. We must require $\hat{\omega} \cdot \vec{r} = 0$. We impose this by substituting a delta function into (2.4) and then integrating over all $\vec{\omega}$.

Thus we have

$$U\Phi(U) = \int \delta(R - |\vec{r}_p|)\delta(\vec{\omega} \cdot \vec{r})\delta(U - |\vec{v}_p|)\Phi_\omega(\vec{\omega})d\vec{r}d\vec{\omega}. \qquad (2.5)$$

We assume all directions of the rod are equally likely and the rods are also equally likely to have any length. So we set
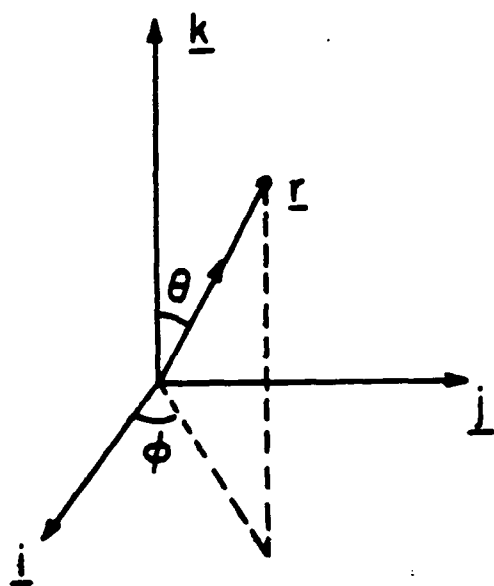
$$\Phi(\vec{r}) = 1. \qquad (2.6)$$

We also assume that the rotation is equally likely to be in any direction. Then $\Phi_\omega(\vec{\omega})$ is a function of $|\vec{\omega}| = \omega$ only.

$$U\Phi(U) = \int \delta(R - |\vec{r}_p|)\delta(\vec{\omega} \cdot \vec{r})\delta(U - |\vec{v}_p|)\Phi_\omega(\omega)d\vec{r}d\vec{\omega}. \qquad (2.7)$$

Choose an orthogonal basis $\vec{i}, \vec{j}, \vec{k}$. Define angles $\theta, \varphi, \alpha, \beta$. These angles are illustrated in figure 1. θ is the angle between the unit vector and the $\vec{k}$ axis. φ is the angle between the unit vector projected onto the x, y plane and the x-axis. Similarly for α and β. It follows that

$$\vec{r} = sin\theta\, cos\varphi\vec{i} + sin\theta\, sin\varphi\vec{j} + cos\theta\vec{k}$$

$$\vec{\omega} = sin\alpha\, cos\beta\vec{i} + sin\alpha\, sin\beta\vec{j} + cos\alpha\vec{k} \qquad (2.8)$$

(i)

(ii)
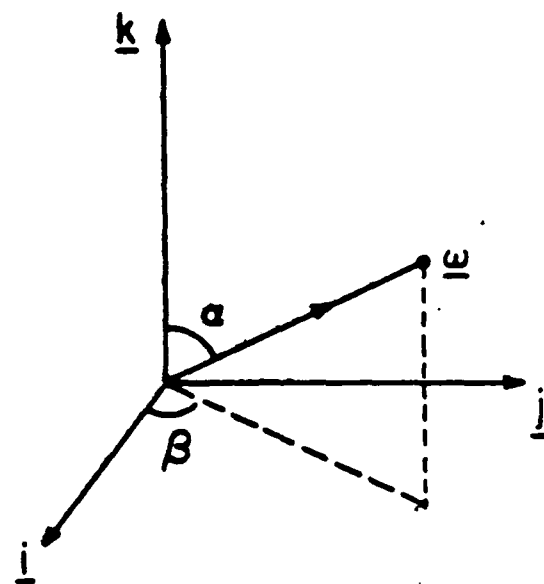
Figure 1 *Definitions of the angles, see text.*

Then

$$d\vec{\omega} = d\omega \, \sin\alpha \, d\alpha d\beta$$

$$\hspace{6cm}(2.9)$$

$$d\vec{r} = dr \, \sin\theta \, d\theta d\varphi$$

and

$$|\vec{r}_p| = r\sin\theta. \hspace{3cm}(2.10)$$

Define the angle between $\vec{\omega}$ and $\vec{r}$ to be τ by

$$cos\tau = \vec{\omega} \cdot \vec{r}. \hspace{3cm}(2.11)$$

Then from (2.8) we have

$$cos\tau = cos\theta\, cos\alpha + sin\theta\, sin\alpha\, cos(\varphi - \beta). \qquad (2.12)$$

Define the angle ρ by

$$cos\rho = \frac{(\vec{\omega} \times \vec{r})}{|\vec{\omega} \times \vec{r}|} \cdot \vec{k}. \qquad (2.13)$$

From (2.11) we obtain

$$cos\rho\, sin\tau = (\vec{\omega} \times \vec{r} \cdot \vec{k}). \qquad (2.14)$$

From (2.8) we find

$$cos\rho\, sin\tau = sin\theta\, sin\alpha(\varphi - \beta) \qquad (2.15)$$

and we calculate

$$|\vec{v}_p| = \omega sin\tau\, sin\rho. \qquad (2.16)$$

We use (2.10), (2.11) and (2.16) to write the delta functions as

$$\delta(R - |\vec{r}_p|)\delta(\vec{\omega} \cdot \vec{r})\delta(U - |\vec{v}_p|) = \delta(R - r\, sin\theta)\delta(cos\tau)\delta(U - \vec{\omega}r\, sin\tau\, sin\rho). \quad (2.17)$$

We now need to change the variables of integration. We have

$$d\vec{\omega}d\vec{r} = d\omega dr \; \sin\alpha \, \sin\theta \; d\alpha d\beta d\theta d\varphi. \qquad (2.18)$$

The integrand only depends on φ and β as $\varphi - \beta$. We can set $\psi = \varphi - \beta$ and integrate out β. Then we have

$$\cos\tau = \cos\theta \, \cos\alpha + \sin\theta \, \sin\alpha \cos\psi$$

$$\cos\rho \sin\tau = \sin\theta \, \sin\alpha \, \sin\psi \cdot \qquad (2.19)$$

$$d\vec{\omega}d\vec{r} = \tau d\omega dr \; \sin\alpha \, \sin\theta \; d\alpha \, d\psi \, d\theta$$

Now we must change the variables of integration from α, ψ to τ, ρ:

$$d\alpha d\psi = \det \begin{pmatrix} \frac{\partial\alpha}{\partial\tau} & \frac{\partial\alpha}{\partial\rho} \\[2mm] \frac{\partial\psi}{\partial\tau} & \frac{\partial\psi}{\partial\rho} \end{pmatrix} d\tau d\rho. \qquad (2.20)$$

From (2.19) we eliminate the terms containing ψ to obtain

$$\cos\alpha = \cos\tau \, \cos\theta \pm \sin\alpha (\sin^2\theta - \cos^2\rho)^{\frac{1}{2}}. \qquad (2.21)$$

Define functions $F(\tau, \rho, \alpha, \psi)$ and $G(\tau, \rho, \alpha, \psi)$ by

$$F(\tau, \rho, \alpha, \psi) = \cos\tau - \cos\theta \, \cos\alpha - \sin\theta \, \sin\alpha \cos\psi$$
$$\qquad (2.22)$$
$$G(\tau, \rho, \alpha, \psi) = \cos\rho \sin\tau - \sin\theta \, \sin\alpha \, \sin\psi$$

It is a standard result of Jacobian transformations (it can be obtained by using formula (2.20) twice, once changing variables from α, ψ to F, G and then changing F, G to τ, ρ) that

$$dad\psi = det \begin{pmatrix} \frac{\partial F}{\partial \tau} & \frac{\partial F}{\partial \rho} \\ \\ \frac{\partial G}{\partial \tau} & \frac{\partial G}{\partial \rho} \end{pmatrix} / det \begin{pmatrix} \frac{\partial F}{\partial \alpha} & \frac{\partial F}{\partial \psi} \\ \\ \frac{\partial G}{\partial \alpha} & \frac{\partial G}{\partial \psi} \end{pmatrix} d\tau d\rho. \qquad (2.23)$$

From (2.22) we obtain

$$det \begin{pmatrix} \frac{\partial F}{\partial \tau} & \frac{\partial F}{\partial \rho} \\ \\ \frac{\partial G}{\partial \tau} & \frac{\partial G}{\partial \rho} \end{pmatrix} = sin^2 \tau \, sin\rho \qquad (2.24)$$

and

$$det \begin{pmatrix} \frac{\partial F}{\partial \alpha} & \frac{\partial F}{\partial \psi} \\ \\ \frac{\partial G}{\partial \alpha} & \frac{\partial G}{\partial \psi} \end{pmatrix} = sin\theta \, sin\alpha \{ sin\theta \, cos\alpha - cos\psi \, cos\theta \, sin\alpha \}. \qquad (2.25)$$

Thus we calculate

$$dad\psi = \frac{sin^2 \tau \, sin\rho}{sin\theta \, sin\alpha \{ sin\theta \, cos\alpha - cos\psi \, cos\theta \, sin\alpha \}} d\tau d\rho. \qquad (2.26)$$

We substitute for $cos\psi$ from (2.19) and obtain

$$dad\psi = \frac{sin^2 \tau \, sin\rho}{sin\alpha \{ cos\alpha - cos\theta \, cos\tau \}} d\tau d\rho \qquad (2.27)$$

then using (2.21) we find

$$dad\psi = \frac{sin\tau \, sin\rho}{sin\alpha (sin^2\theta - cos^2\rho)^{\frac{1}{2}}} d\tau d\rho. \qquad (2.28)$$

Hence

$$sin\alpha \, sin\theta dad\psi = \frac{sin\tau \, sin\rho \, sin\theta}{(sin^2\theta - cos^2\rho)^{\frac{1}{2}}} d\tau d\rho. \qquad (2.29)$$

71

We substitute (2.17), (2.18) and (2.29) into (2.7)

$$U\Phi(u) = \int \delta(R - r\sin\theta)\delta(\cos\tau)\delta(u - \omega r\sin\tau \sin\rho)$$

$$\Phi_\omega(\omega)\frac{\sin\tau \sin\rho \sin\theta}{(\sin^2\theta - \cos^2\rho)^{\frac{1}{2}}}\, d\tau\, d\rho\, d\theta\, d\omega\, dr \qquad (2.30)$$

Now we do the integration with respect to τ to obtain

$$U\Phi(U) = \int \frac{\delta(R - r\sin\theta)\delta(U - \omega r\sin\rho)}{(\sin^2\theta - \cos^2\rho)^{\frac{1}{2}}}\Phi_\omega(\omega)\sin\rho \sin\theta\, d\rho\, d\theta\, d\omega\, dr. \qquad (2.31)$$

We integrate with respect to θ to obtain

$$U\Phi(u) = \int \frac{\delta(U - \omega r\sin\rho)\Phi_\omega(\omega)\sin\rho(R/r)\, d\rho\, d\omega\, dr}{(r^2 - R^2)^{\frac{1}{2}}(R^2/r^2 - \cos^2\rho)^{\frac{1}{2}}}. \qquad (2.32)$$

We integrate with respect to ρ and obtain

$$\Phi(U) = \int \frac{R\Phi_\omega(\omega)d\omega dr}{r(r^2 - R^2)^{\frac{1}{2}}(R^2\omega^2 + u^2 - \omega^2 r^2)^{\frac{1}{2}}(\omega^2 r^2 - u^2)^{\frac{1}{2}}}. \qquad (2.33)$$

The domain of integration is restricted to the region of the $\omega - r$ plane where

the integrand is real (for simplicity we have omitted the bounds of integration

in the previous equations). It is specified by

$$r \geq R$$

$$R^2\omega^2 + u^2 \geq \omega^2 r^2 \qquad (2.34)$$

$$\omega r \geq u.$$

Change the variable of integration from r to X where

$$\omega^2 r^2 = u^2 + X \qquad (2.35)$$

then

$$\Phi(u) = \int \frac{R\omega \Phi_\omega(\omega) d\omega dX}{2X^{\frac{1}{2}}(R^2\omega^2 - X)^{\frac{1}{2}}(u^2 + X - \omega^2 R^2)^{\frac{1}{2}}(u^2 + X)} \qquad (2.36)$$

with domain of integration

$$X \le R^2\omega^2$$

$$X \ge 0 \cdot \qquad (2.37)$$

$$u^2 + X \ge \omega^2 R^2$$

Define a new variable Y by

$$Y^2 = \omega^2 R^2 - X \qquad (2.38)$$

then

$$\Phi(u) = \int \frac{R\omega \Phi_\omega(\omega) d\omega dY}{(\omega^2 R^2 - Y^2)^{\frac{1}{2}}(u^2 - Y^2)^{\frac{1}{2}}(u^2 + \omega^2 R^2 - Y^2)} \qquad (2.39)$$

with

$$Y \geq 0$$

$$Y \leq \omega R \tag{2.40}$$

$$Y \leq u.$$

We can further simplify this expression by the substitutions $Y = u \sin\theta$ and $\omega = (u/R)\Omega$. This yields

$$\Phi(u) = \frac{1}{u^2} \int_0^{\frac{\pi}{2}} d\theta \int_{\sin\theta}^{\infty} d\Omega \frac{\Phi_\omega(u\Omega/R)}{(\Omega^2 - \sin^2\theta)^{1/2}(\Omega^2 + \cos^2\theta)}. \tag{2.41}$$

It is clear from this expression that $\Phi(u)$ will decrease with u if and only if $\Phi_\omega(\omega)$ does not grow faster than ω^2. Thus for all realistic cases $\Phi(u)$ will be a monotonic decreasing function of u.

Section 3. Expansion and Contraction.

We now consider the situation in two dimensions when the rod is allowed to expand and contract as well as rotate. We can write

$$\vec{r} = r\vec{\hat{r}} \tag{3.1}$$

and differentiating gives

$$\dot{\vec{r}} = \dot{r}\vec{\hat{r}} + r\dot{\theta}\vec{\hat{n}} \tag{3.2}$$

where $\vec{\hat{n}}$ is the unit vector perpendicular to $\vec{\hat{r}}$. We define the expansion coefficient s by

$$\dot{r} = rs \tag{3.3}$$

and write (3.2) as

$$\vec{\dot{r}} = rs\vec{\hat{r}} + r\dot{\theta}\vec{\hat{n}}. \tag{3.4}$$

The projected velocity is then given by

$$U = r\,s\,sin\theta + r\,\omega\,cos\theta \tag{3.5}$$

where θ is the angle that $\vec{\hat{r}}$ makes with the normal to the image plane and $\dot{\theta} = \omega$. The projected velocity distribution is given by

$$\Phi(U) = \int \delta(R - rsin\theta)\delta(U - (r\,s\,sin\theta + r\omega cos\theta))\Phi_s(s)\Phi_\omega(\omega)\,dr\,d\omega\,ds\,d\theta$$
$$\tag{3.6}$$

where, as in Section 1, we have assumed that $\Phi(r) = 1$ so the rods have no preferred lengths. We do the integral of $\Phi(U)$ with respect to θ to obtain

$$\Phi(U) = \int (r^2 - R^2)^{-1/2}\delta(U - sR - \omega(r^2 - R^2)^{1/2})\Phi_s(s)\Phi_\omega(\omega)\,dr\,d\omega\,ds. \tag{3.7}$$

We now integrate with respect to r to find

$$\Phi(U) = \int ((U - sR)^2 + \omega^2 R^2)^{-1/2}\Phi_s(s)\Phi_\omega(\omega)d\omega ds. \tag{3.8}$$

Care must be taken to ensure that this integral is evaluated over the correct limits, those for which the integrand is well defined. For $\omega > 0$ we need $s < \frac{U}{R}$ and for $\omega < 0$ we have $s > \frac{U}{R}$. This domain of integration is shown in figure (2). We can split the integral up into two parts

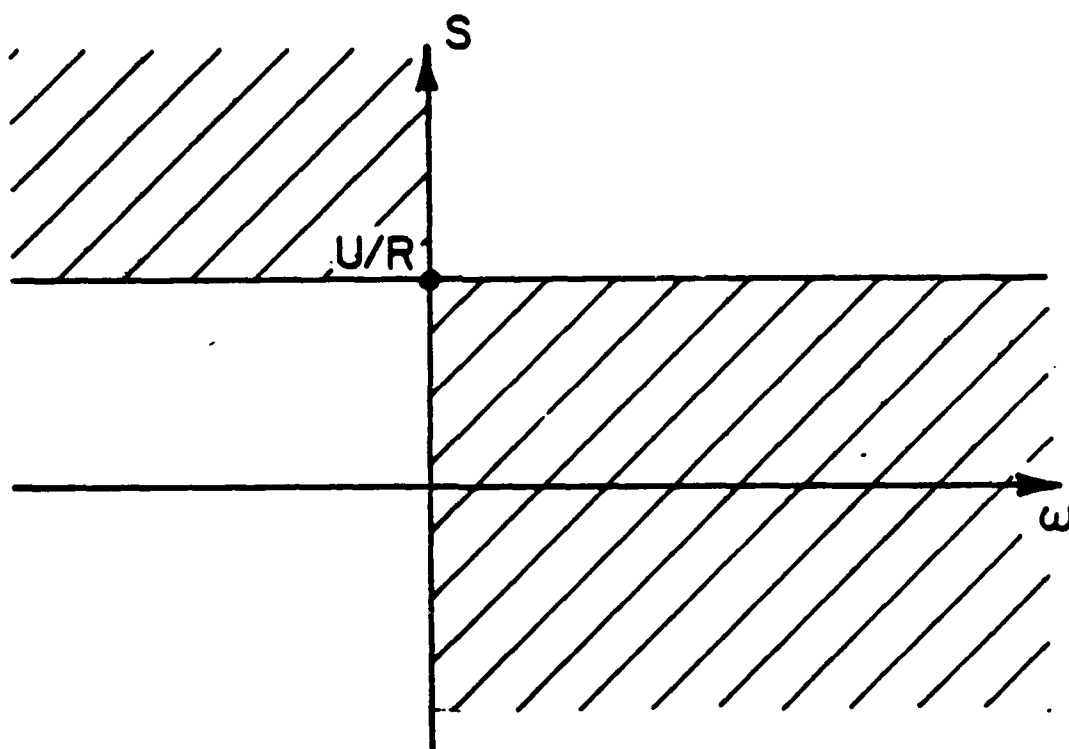$$\Phi(U) = \Phi_1(U) + \Phi_2(U) \tag{3.9}$$

Figure 2 *The domain of integration, see text.*

where

$$\Phi_1(U) = \int_{\omega=0}^{\omega=\infty} d\omega \int_{s=-\infty}^{s=U/R} ((U - sR)^2 + R^2\omega^2)^{-1/2} \Phi_s(s)\Phi_\omega(\omega)ds \tag{3.10}$$

and

$$\Phi_2(U) = \int_{\omega=-\infty}^{\omega=0} d\omega \int_{s=U/R}^{s=\infty} ((U - sR)^2 + R^2\omega^2)^{-1/2} \Phi_s(s)\Phi_\omega(\omega)ds. \quad (3.11)$$

We impose the conditions

$$\Phi_s(s) = \Phi_s(-s) \quad (3.12)$$

and

$$\Phi_\omega(\omega) = \Phi_\omega(-\omega). \quad (3.13)$$

These ensure that contraction is as likely as expansion and that rotation is equally likely to be any direction.

We use (3.12) and (3.13) to rewrite (3.9) as

$$\Phi(U) = \int_{\omega=0}^{\omega=\infty} d\omega \int_{s=-\infty}^{s=U/R} ((U - sR)^2 + R^2\omega^2)^{-1/2} \Phi_s(s)\Phi_\omega(\omega)ds$$

$$+ \int_{\omega=0}^{\omega=\infty} d\omega \int_{s=-\infty}^{s=-U/R} ((U + sR)^2 + R^2\omega^2)^{-1/2} \Phi_s(s)\Phi_\omega(\omega)ds. \quad (3.14a)$$

We now substitute $t = Rs/u$ and $\omega = u\Omega/R$ and obtain

$$\Phi(U) = \frac{u}{R^2} \int_{\Omega=0}^{\Omega=\infty} d\Omega \int_{t=-\infty}^{t=1} ((1 - t)^2 + \Omega^2)^{-1/2} \Phi_s(ut/R)\Phi_\omega(u\Omega/R)dt$$

$$+ \frac{u}{R^2} \int_{\Omega=0}^{\Omega=\infty} d\Omega \int_{t=-\infty}^{t=-1} ((1 + t)^2 + \Omega^2)^{-1/2} \Phi_s(ut/R)\Phi_\omega(u\Omega/R)dt. \quad (3.14b)$$

Note that if we set $\Phi_s(ut/R) = \delta(ut/R)$ we recover (1.3).

We differentiate (3.14a) to obtain

$$\frac{\partial \Phi}{\partial U} = -\int_{\omega=0}^{\omega=\infty} d\omega \int_{s=-\infty}^{s=U/R} ((U+sR)^2 + R^2\omega^2)^{-3/2}(U+sR)\Phi_s(s)\Phi_\omega(\omega)ds$$

$$-\int_{\omega=0}^{\omega=\infty} d\omega \int_{s=-\infty}^{s=-U/R} ((U-sR)^2 + R^2\omega^2)^{-3/2}(U-sR)\Phi_s(s)\Phi_\omega(\omega)ds. \quad (3.15)$$

We can simplify this expression by setting s to $s - 2U/R$ in the second term

on the right hand side. This yields

$$\frac{\partial \Phi}{\partial U} = -\int_{\omega=0}^{\omega=\infty} d\omega \int_{s=-\infty}^{s=-U/R} \frac{1}{((U+sR)^2 + R^2\omega^2)^{3/2}}(sR+U)(\Phi_s(s) - \Phi_s(s+2U/R))\Phi_\omega(\omega)ds.$$

$$(3.16)$$

In the range of integration $(u + sR)$ is always negative. So the integral will

be always negative provided $\Phi_s(s) - \Phi_s(s + 2u/R)$ is negative for $s \le -u/R$.

If we make s positive, using (3.12), this becomes

$$\Phi_s(s - 2U/R) - \Phi_s(s) \ge 0, \qquad (3.17)$$

which is true for all monotonically decreasing functions. Then we have $\frac{\partial \Phi}{\partial U} < 0$

everywhere for any function $\Phi_\omega(\omega)$. Hence the projected velocity distribution

will once more be peaked about $U = 0$.

If we relax the conditions (3.17) we are no longer assured of having a

monotonically decreasing projected velocity distribution for all $\Phi_\omega(\omega)$. If

$\Phi_s(s)$ is linear in s we obtain a flat distribution

$$\Phi(U) = const. \tag{3.18}$$

If $\Phi_s(s)$ is quadratic in s we find

$$\frac{\partial \Phi}{\partial U} = -2 \int d\omega \frac{\Phi_\omega(\omega)}{\omega} \frac{U^2}{R^4} + \int \int d\omega ds s^2 (s^2 + \omega^2)^{-3/2} \frac{4U}{R^3} \Phi_\omega(\omega) \tag{3.19}$$

which will only be negative for some distributions $\Phi_\omega(\omega)$.

Thus allowing the rod to expand and contract reduces the strength of the result of Section 1. But if we make the reasonable assumption that $\Phi_s(s)$ is at most linear in s the result will still hold.

Section 4. Curves from rods.

The situation becomes more complex when we consider several rods joined together to form a curve. The results above show that for each rod individually the projected velocity distribution is peaked at zero. If the rods are joined, however, these probability distributions are no longer independent. We now sketch an argument suggesting that this loss of independence does not affect the result.

Suppose we have a set of probability distributions $P_i(x_i)$, $i = 1, ..N$ which are all peaked at zero. We now define the joint probability distribution $P(\vec{x})$ where $\vec{x} = (x_i, x_2, ..., x_N)$. If we impose no constraints this distribution is given by

$$P(\bar{x})d\bar{x} = \Pi_i P_i(x_i)\Pi_i dx_i. \tag{4.1}$$

Now impose a consistency constraint $\sum_i x_i = 0$. We now have

$$\int P(\bar{x})d\bar{x} = \int \Pi_i P_i(x_i)\delta(\sum_i x_i)\Pi_i dx_i. \tag{4.2}$$

The x_i's can no longer vary independently and the delta function imposes the consistency constraint. We can get integrate out the delta function thereby reducing the number of independent variable to $N-1$. It is convenient to choose these to be $x_i - x_{i+1}$, for $i = 1, ..., N-1$. We can now write (4.2) as

$$\int P(\bar{x})d\bar{x} = \int \Pi_i P_i(x_i - x_{i+1})\Pi_i d(x_i - x_{i+1}). \tag{4.3}$$

We can set $y_i = x_i - x_{i+1}, \quad i = 1, ..., N-1$ and rewrite

$$P(\bar{y}) = P_1(y_1)P_2(y_2)...P_{N-1}(y_{N-1})P_N(-y_1 - y_2 - ... - y_{N-1}). \tag{4.4}$$

It is straightforward to see that the maximum value of $P(\bar{y})$ occurs when the y_i are all zero. Thus P is still peaked at zero even with the constraints.

Section 5. An alternative approach.

The results derived in the last few sections support the assumption (Hildreth 1984) that the velocity field along a contour is smooth. This assumption

is used to solve the aperture problem by prescribing the smoothest possible velocity field consistent with the data.

An alternative approach to the aperture problem was taken by Waxman and Wohn (1986). They showed that if the object was locally planar and moving rigidly then, assuming perspective projection, the local velocity field on the image plane would be locally quadratic in the x, y coordinates. They then defined regions in which these expansions were valid and checked for consistency between these regions. In these regions they could do a least squares best fit to find the coefficients of these quadratic polonomials, and hence the velocity field.

This approach, however, can also be used to justify the smoothest velocity field assumption. The orthographic projection of the motion of a rigid body will obey

$$v_x = A + By + Cz, \qquad\qquad (5.1a)$$

$$v_y = D + Ex + Fz, \qquad\qquad (5.1b)$$

where A, B, C, D, E, F are constants and z is the (unknown) depth coordinate. If we also assume the object is planar then z obeys

$$z = px + qy + s, \tag{5.2}$$

where p, q, r are constants. Substituting (5.2) into (5.1) gives linear expressions for v_x and v_y in terms of x and y. So if we assume local planarity and local rigidity (in the style of Waxman and Wohn (1986)) the velocity fields will locally be linear in x and y. The measure of smoothest velocity, $I(\vec{v})$, used by Hildreth is the integral along the curve of the function

$$J(\vec{v}) = \frac{\partial \vec{v}}{\partial s} \cdot \frac{\partial \vec{v}}{\partial s}, \tag{5.3}$$

where s is the arc length. The velocity fields will be locally linear if and only if $J(\vec{v})$ is zero. Thus we can think of the smoothest velocity field approach as a local method of assuming local rigidity and local planarity.

Sectio ι 6. Conclusion.

The arguments in the first four sections of this paper suggest that locally rigid, or semi-rigid objects, will tend to project a smooth velocity field in the image. Moreover, assuming random motions and limited expansion or contraction, this field will tend to be as smooth as possible. In the fifth section we noted that rigidity of an object and local smoothness of its surface will also lead to a smooth image motion.

Tnese arguments support the view that maximixing smoothness is a good heuristic to use for motion correspondence and that it is a sensible way to solve the aperture problem.

Acknowledgements

We would like to thank Ellen Hildreth and Tomaso Poggio for helpful comments on this manuscript.

Appendix

We first describe the general method for integrating delta functions.

Suppose we have an integral $I(a,b)$

$$I(a,b) = \int_a^b f(x)\delta(x - x_0)dx \qquad (A.1)$$

where $\delta(x)$ is the Dirac delta function, $f(x)$ is an arbitrary function, x_0 an arbitrary point and $b \geq a$. The value of the integral is

$$I(a,b) = f(x_0), \quad if \quad x_0 \epsilon [a,b] \qquad (A.2a)$$

$$I(a,b) = 0, \quad otherwise. \qquad (A.2b)$$

This result can be generalized to integrals of form

$$J(a,b) = \int_a^b f(x)\delta(g(x) - c)dx \qquad (A.3)$$

where $g(x)$ is an arbitrary function and c an arbitrary number. All points x_i with $g(x_i) = c$ will contribute to this interval. Suppose there is only one such point. If there are several we can divide the integral up into regions with only one such point. Consider one such point $x = 0$. The function $g(x)$ can be expanded in a Taylor series about this point

$$g(x) = c + g'(0)(x) + O(x^2). \qquad (A.4)$$

If we change the coordinate to u where $u = g'(0)x$ we can write the integral as

$$J(a,b) = \int_{g'(0)a}^{g'(0)b} f(u/g'(0))\delta(u + O(u^2))\frac{du}{g'(0)}. \qquad (A.5)$$

The value of the integral will depend on the sign of $g'(0)$. If it is negative the bounds of the integral will be reversed and the integral will change sign. Therefore

$$J(a,b) = f(0)\frac{1}{|g'(0)|}, \quad 0\epsilon[a,b] \qquad (A.6a)$$

$$J(a,b) = 0, \quad otherwise. \qquad (A.6b)$$

We now consider the form of the probability distribution function for a rotational symmetric vector in two-dimensional space. Suppose the function is $\Phi_{\vec{\omega}}(\vec{\omega})d\vec{\omega}$. If it is rotationally symmetric (and hence depends only on the modulus ω of $\vec{\omega}$) it can be written

$$\Phi_{\vec{\omega}}(\vec{\omega})d\vec{\omega} = \Phi_{\omega}(\omega)\omega d\omega d\varphi, \qquad (A.7)$$

by changing to radial coordinates ω, φ in the ω space. The φ component can be integrated out to give a distribution $\omega\Phi_{\omega}(\omega)d\omega$.

References

Fennema, C.L. and Thompson, W.B. 1979. Velocity determination in scene containing several moving objects,*Comp. Graph. Im. Proc. (9)* 301-315

Hildreth, E.C. 1984. *The Measurement of Visual Motion.* Cambridge: MIT Press.

Horn, B.K.P. and Schunck, B.G. 1981. "Determining optical flow." *Art. Intell. (17)* 185-203.

Nakayama, K. and Silverman, G.H. 1986. "The aperture problem II: Spatial integration of velocity information along contours." *Submitted for publi-*

cation

Ullman, S. 1979. *The Interpretation of Visual Motion*. Cambridge: MIT Press.

Ullman, S. 1984. "Maximizing rigidity: the incremental recovery of 3-D structure from rigid and non-rigid motion," it Perception (13), 255-274.

Waxman, A.M. and Wohn, K. 1986. "Image flow theory: A framework for 3-D inference from time-varying imagery," In *Advances in Computer Vision*. Eds C.Brown. Erlbaum Publishers.

7.0 Assessment and Future Directions

Computer Graphics provides a powerful tool for understanding vision. Although it is exceedingly time consuming to generate realistic models of a scene *from physical principles*, the rewards are great. There is no other method available at present of comparable power that allows us to isolate the physical parameters relevant to "seeing" natural surfaces or objects. Renditions now exist for many classes of plants and trees, and for several kinds of natural phenomena like water, clouds and sunsets (see SIGGRAPH and Fournier & Reeves, 1987). These programs present an unusual opportunity for the visual psychophysicist who wishes to understand the perception of natural surfaces, and how inferences about material types and objects can be made from images.

One way to further the Graphics Psychophysics approach would be to encourage psychophysicists to collaborate with those laboratories actively engaged in the generation of natural phenomena using graphics. Typically these graphics labs are interested in selecting the best rendition parameters, but do not bother to conduct parametric studies which the psychophysicist is trained to do. Graphics laboratories at Cornell (Architecture), University of North Carolina, Toronto, Cal. Tech., NYIT and MIT-Media might be approachable hosts. A special sabbatical or post-doc fellowship to a young investigator would encourage this collaboration at little cost. Psychophysical studies of the rendition of even a half-dozen different materials would greatly advance image understanding.

Finally, this project only began to explore time-dependent phenomena, namely the motion and creation of water waves, plus some semi-rigid motions. This relatively unexplored area of non-rigid motions should be contrasted with the large number of studies which are devoted to rigid body motions or to flow fields. Non-rigid motions include not only the wave motions of fluids such as water, honey, milk, oil, etc. (which we can tell apart psychophysically), but also turbulent motions, deformations, and articulated motions. Here we also have lacunae well worth filling.

Appendix I: Relating Fractal Surfaces to Fractal Images

Our aim is to recover a 3D fractal measure of surface roughness. As a first step, we need to show that the resultant image contains a measure related to the 3D fractal specification for the surface. Pentland (1983) and Kube & Pentland (1986) have already shown that a convex Lambertian 3D surface of constant albedo with a spatially isotropic fractal Brownian shape seen under constant illumination produces an image whose intensity surface is also fractal Brownian with a fractal dimension identical to that of the components of the surface normal. Specifically, given

$$I(x,y) = \rho E(N \cdot L) \tag{1}$$

where ρ, E, L are constant and N is a Brownian fractal, then $I(x,y)$ will be fractal.

We wish to prove a similar result for glossy surfaces. In particular, we require that I will obey the rule

$$Pr\left(\frac{I(x,y) - I(x + \Delta x, y)}{\|\Delta x\|^h} < u \right) = F(u) \tag{2}$$

where h is the Hausdorf dimension and

$$I(x,y) = E \cdot (N \cdot H) \cdot F_{fresnel} \tag{3}$$

where $H = L + V$ (i.e. bisector of illuminant and viewer directions). $F_{fresnel}$ is the Fresnel reflectance function. Unfortunately, $F_{fresnel}$ is dependent upon the relation between N and L (incident angle), even when N and H align, which is seldom for point sources. Together, these factors generally will prohibit a point source from generating specularity regions that satisfy a fractal image statistic.

However, most scenes are illuminated also by diffuse, hemispheric sources. Under these conditions, disregarding occlusions, for every microfacet viewed there will be a patch of the hemisphere reflected off the facet. Equation (3) now becomes

$$I(x,y) = E^* \cdot F_{fresnel}(N \cdot V) \tag{4}$$

where the Fresnel term is a function only of the emittant angle, which equals the incident angle, and E^* is illumination per solid angle.

We next approximate the Fresnel reflectance by

$$F = (1 - V \cdot N)^n = (1 - \cos\theta)^n \tag{5}$$

where n is a parameter related to the index of refraction, and controls the rate at which the Fresnel reflectance falls toward zero as the light beam moves away from its maximum reflectance of 1.0 for the grazing condition. For incident angles θ greater than $60°$ we can approximate (5) by its Taylor Series expansion, retaining only the first terms:

$$F = 1 - nV \cdot N \tag{6}$$

Equation (6) is then substituted into (4), which in turn can be used to test for condition (1). Dividing out the common factor E^*, we find that

$$Pr\left(\frac{1 - nN \cdot V - 1 + n(N + \Delta N) \cdot V}{\|\Delta x\|^h} < u\right) = F(u) \tag{7}$$

or

$$Pr\left(\frac{-N \cdot V + (N + \Delta N) \cdot V}{\|\Delta x\|^h} < \frac{u}{n}\right) = F(u) \quad . \tag{8}$$

But as long as V is a constant, which it is for viewing glossy surfaces at sufficient distances, then the fractal nature of N will result in a scaled version of (8) satisfying condition (1). Hence glossy reflections off a fractal 3D surface seen under hemifield illumination will produce a fractal image-intensity surface.

Appendix II Recovering Material Properties from Sound (R.P. Wildes)

1.0 Introduction

Physical modeling of world events occupies an important position in computational approaches to perception and motor control (Brady et al., 1982; Horn, 1970; Marr, 1982; Richards, 1988). Such models are dual in purpose. First, they lend to a precise statement of the problem under consideration. Second, they expose constraints on the solution space, which indicate how the problem can be solved. Here, we use this approach for an initial attack on a previously unsolved problem in audition: How it is possible to recover the material property of an object from the sound generated when it is struck?

Consider some examples. When you hit a drinking glass with your finger-nail it produces a distinct "ringing" sound. In comparison, if you strike a log with your knuckles it gives off a short "thud". In either case you easily identify the material of the struck object. There are other cases where the sound from striking an object leaves one with no clear impression of material type. Objects which are clamped or otherwise artificially damped often belong to this latter class. Our goal is then two-fold: First, we seek to discover a physical parameter of the sound following impact which is intrinsically related to material type. Second, we desire to recognize those situations where such material identification is not possible. Guided by a model of vibrating solids we shall choose to satisfy these goals with a measure of energy dissipation during vibration—an intrinsic material property.

2.0 A Physical Model

When a solid object is subject to an impact, much of the resultant sound is due to the vibration of the object. Therefore, we study the mechanics of vibrating solids, and take that motion as analogous to the sound production. In the following presentation we do not consider the initial transient due to impact. Instead, we concentrate on the steady-state and damped behavior of the solid. We proceed in two stages. First, we introduce some basic concepts concerning the deformation of solids. Second, we develop a particular model of a vibrating solid: the standard anelastic linear solid described by Zener (1948).[1]

---

[1] Many of the results in Sections 2.1, 2.2 and 3.1 are known in the literature on anelasticity. For further details see Nowick & Berry, 1972; Wert, 1986; Zener, 1948.

2.1 Basic Concepts

Hooke's law states that an ideal elastic material can be described by the relation

$$\sigma = M\epsilon \tag{1}$$

Where: σ is a stress variable (corresponding to a force); ϵ is a strain variable (corresponding to a displacement); and M is an appropriate modulus of elasticity.[2] In many situations it is convenient to define $J = \frac{1}{M}$, as the modulus of compliance. Then (1) becomes

$$\epsilon = J\sigma \tag{2}$$

Elastic behavior so defined is seen to be characterized by three conditions. First, there is a $1:1$ correspondence between stress and strain. Second, the stress-strain relationship is linear. Third, there is no time dependence in the mapping between stress and strain. Clearly, any real solid is only approximately characterized by these conditions. In most cases, condition three is a particularly erroneous assumption. Therefore, we proceed by considering cases where there is a time dependence between stress and strain. Solid materials characterized by such a time dependence are called anelastic solids.

Suppose a specimen is subjected to a periodic stress

$$\sigma = \sigma_0 e^{i\omega t} \tag{3}$$

Here: σ_0 is the stress amplitude; $\omega = 2\pi f$ is the angular frequency; t is time; and $i = \sqrt{-1}$. Then, our assumptions of linearity and time dependence dictate that strain has the form

$$\epsilon = \epsilon_0 e^{i(\omega t - \phi)} \tag{4}$$

with ϵ_0 the strain amplitude and ϕ a phase angle. Then, ϕ captures the notion of a time dependence between stress and strain. For subsequent operations it is convenient to separate (4) into its real and imaginary parts. Letting ϵ_1 be the component of strain in phase with stress and ϵ_2 be that component of strain $\frac{\pi}{2}$ out of phase with stress we have

$$\epsilon = (\epsilon_1 - i\epsilon_2)e^{i\omega t} \tag{5}$$

Now, in analogy with (2) we define the complex compliance

$$J(\omega) = \frac{\epsilon}{\sigma} \tag{6}$$

Then, dividing (4) by (3) gives

---

[2]Notice that for a fine grained analysis σ and ϵ would be second order tensors. For our purposes it suffices to consider them as scalars.

$$J(\omega) = \frac{\epsilon_0}{\sigma_0} e^{i\phi(\omega)} \tag{7}$$

The complex compliance can also be represented as a sum of real and imaginary parts by dividing (5) by (3) and defining

$$J(\omega) = J_1(\omega) - iJ_2(\omega) \tag{8}$$

We proceed by expanding (7) in terms of Euler's relation[3] and then equating real and imaginary parts of (7) and (8) to get

$$J_1(\omega) = \frac{\epsilon_0}{\sigma_0} \cos \phi(\omega)$$
$$J_2(\omega) = \frac{\epsilon_0}{\sigma_0} \sin \phi(\omega) \tag{9}$$

From which we reach the important result that

$$\tan \phi = \frac{J_2}{J_1} \tag{10}$$

In the anelasticity literature $\tan \phi$ is referred to as internal friction. Internal friction is an intrinsic property of a given material; it measures the degree of anelasticity. We shall ultimately recover a measure closely related to internal friction as our characterization of solid materials.

Finally, it is worth noting that analogous calculations can be carried out in terms of the modulus of elasticity. Such derivations would begin by defining

$$M(\omega) = \frac{\sigma}{\epsilon} \tag{11}$$

with the ultimate conclusion that

$$\tan \phi = \frac{M_2}{M_1} \tag{12}$$

2.2   The Standard Anelastic Linear Solid

The model of solids that we wish to consider is depicted as a mechanical structure in Figure 1. We refer to this model as the standard anelastic linear solid. The model consists of three elements. Elements a and b are Hookean springs; let them have corresponding compliances J_a and J_b. Element c is a Newtonian dashpot, which serves to provide internal friction. A Newtonian dashpot can be thought of as a plunger moving in a viscous liquid. Its velocity of motion is proportional to the applied force. In terms of stress and strain

$$\sigma = \eta \dot{\epsilon} \tag{13}$$

---

[3] $e^{i\theta} = \cos \theta + i \sin \theta$
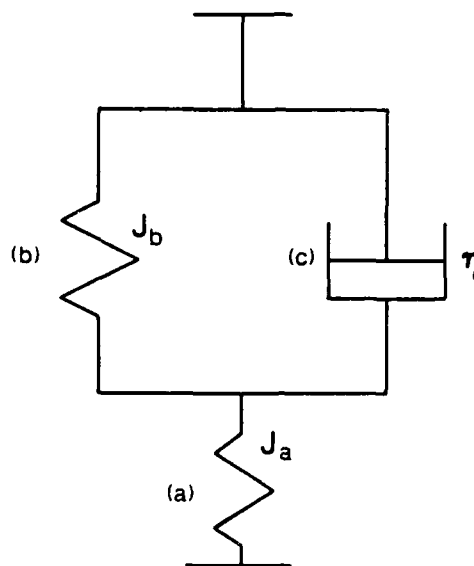
Figure 1. A mechanical model of the standard anelastic linear solid. J_a and J_b are compliances of the two springs. η is the viscosity of the "liquid" in the dashpot

with η the viscosity of the liquid. For our purposes we let $\eta = \frac{\tau_\sigma}{J_b}$, where τ_σ is the time required for the system to achieve equilibrium when a constant stress is applied. Notice that this model captures our intuitive notion of what happens when a force is applied to a solid: Initially the solid deforms (element a). With time further deformation takes place (element c). When the force is removed the solid returns to normal (elements a and b).[4]

In order to be more precise, we characterize the behavior of our model in terms of the differential equation

$$a_0\sigma + a_1\dot{\sigma} = b_0\epsilon + b_1\dot{\epsilon} \tag{14}$$

where the derivatives are with respect to time. We begin by setting

$$\epsilon_a = J_a\sigma_a$$
$$\epsilon_b = J_b\sigma_b \tag{15}$$
$$\dot{\epsilon}_c = J_b\frac{\sigma_c}{\tau_\sigma}$$

---

[4]In some cases the solid never returns to its original configuration. This introduces the concept of visco-elasticity; we shall not consider such cases.
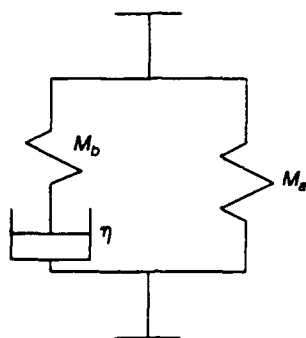
Figure 2 A second mechanical model for the standard anelastic linear solid.

We now introduce the standard rules for combining multiple stresses and strains: For series combination, stresses are equal while strains add. The reverse is true for parallel combinations. Applying these rules to the model of Figure 1 gives

$$
\begin{aligned}
\epsilon &= \epsilon_a + \epsilon_b \\
\epsilon_b &= \epsilon_c \\
\sigma &= \sigma_a \\
&= \frac{\epsilon_a}{J_a} \\
&= \sigma_b + \sigma_c \\
&= \frac{\epsilon_b + \tau_\sigma \dot{\epsilon}_c}{J_b}
\end{aligned}
\tag{16}
$$

We can rearrange (16) to the form (14) as

$$
J\sigma + \tau_\sigma J_a \dot{\sigma} = \epsilon + \tau \dot{\epsilon}
\tag{17}
$$

where $J = J_a + J_b$.

We now characterize our model in terms of the standard measure of internal friction $\tan\phi$. Following methods similar to those of Section 2.1, we substitute (5) and (5) into (17) and equate real and imaginary parts to find that

$$
\begin{aligned}
J_1 &= J_a + \frac{J_b}{1 + (\omega\tau_\sigma)^2} \\
J_2 &= J_b \frac{\omega\tau_\sigma}{1 + (\omega\tau_\sigma)^2}
\end{aligned}
\tag{18}
$$

Then, upon substituting into (10) we get

$$
\tan\phi = J_b \frac{\omega\tau_\sigma}{J + J_a(\omega\tau_\sigma)^2}
\tag{19}
$$

We continue by deriving a more convenient form for $\tan \phi$. Consider the model of Figure 2. This model can be shown to be formally equivalent to the model of Figure 1. It is customary to represent this latter model in terms of the elastic moduli and the time to strain equilibrium τ_ϵ. Following the methods used above and using analogous symbols, we can find that

$$M_1 = \frac{M_a + M_b(\omega\tau_\epsilon)^2}{1 + (\omega\tau_\epsilon)^2}$$

$$M_2 = M_b \frac{\omega\tau_\epsilon}{1 + (\omega\tau_\epsilon)^2} \tag{20}$$

Finally, to get our simpler expression for $\tan \phi$ we combine (18), (19) and (20) to arrive at

$$\tan \phi = \frac{J_b}{(J_a J)^2} \frac{\omega\tau}{1 + (\omega\tau)^2} \tag{21}$$

with τ the geometric mean of τ_σ and τ_ϵ.

We conclude this section with a recapitulation. In this section we began by introducing the concepts of stress, strain and internal friction. We proceeded to present the standard linear model of anelastic solids in terms of these three parameters. In Section 3 we make use of these relations to classify solid materials from the sound they produce following impact.

3.0 Recovery of Material Type

When an anelastic solid is struck and set into vibration its behavior, and hence the sound it generates following a brief transient, is dictated by our standard linear model. Our goal is to extract some intrinsic parameter of the solid material from this dynamic behavior. As noted earlier, internal friction, $\tan \phi$, is just such a measure. We now proceed to derive two measures of internal friction and then show how they can work together.

3.1 Internal Friction and Bandwidth, Q^{-1}

We begin by relating internal friction to peak vibration of our standard linear model. To facilitate this set of derivations, we follow (Zener, 1948) and momentarily consider the mechanical system as having only one degree-of-freedom. This is depicted in Figure 3. Here we have lumped the inertia of the system into a single member I, while the anelastic and elastic components are lumped into an anelastic spring with complex modulus[5]

---

[5] It can be shown that an arbitrary standard anelastic linear system in vibration can always be so reduced, see (Nowick & Berry, 1972, appendix A.)
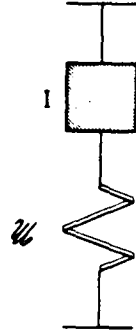
Figure 3 A lumped model of Figures 1 and 2.

$$M = M(1 + i\tan\phi) \tag{22}$$

From Newton's law we observe that

$$I\ddot{\epsilon} = \sigma \tag{23}$$

Assuming periodic motion, we also have that

$$\epsilon = \epsilon_0 e^{i\omega t} \tag{24}$$

and as always

$$\sigma = M\epsilon \tag{25}$$

We can solve for ϵ by substituting (22), (24), and (25) into (23) to get

$$\epsilon = \frac{\frac{\sigma}{m}}{(1 - \frac{I\omega^2}{M}) + i\tan\phi} \tag{26}$$

By inspection, we find (26) at a maximum when

$$\omega = \omega_0 = \sqrt{\frac{M}{I}} \tag{27}$$

Further consideration shows that (26) is at $2^{-\frac{1}{2}}$ maximum amplitude when

$$\omega^2 = \omega_{\pm\frac{1}{2}}^2 = \frac{M}{I}(1 \pm \tan\phi) \tag{28}$$

We arrive at our first measure of internal friction by noting that

$$\frac{\omega_{+\frac{1}{2}} - \omega_{-\frac{1}{2}}}{\omega_0} = Q^{-1} = \tan\phi \tag{29}$$

where we use Q in accordance with the corresponding concept from electrical circuit theory. Thus, internal friction can be measured in terms of the sharpness of the peak around the maximum of vibration of our specimen. This will correspond to a sharpness in the peak of the acoustic signal generated by this vibration. It is this acoustic peak which we suggest measuring.

3.2 Internal Friction and Decay Rate t_e

We now seek a measure of internal friction in terms of the decay in amplitude of a vibrating anelastic solid. Combining (22), (23) and (25) we find that

$$I\ddot{\epsilon} + M(1 + i\tan\phi)\epsilon = 0 \tag{30}$$

describes the behavior of our standard linear system as the vibration decays. The solution to (30) is found to be

$$\epsilon = \epsilon_0 e^{-\Delta ft} e^{i\omega t} \tag{31}$$

with $\Delta = ln(A_n/A_{n+1})$, where A_n and A_{n+1} are successive peaks in the amplitude of the decaying tone of frequency f. (Δ is called the logarithmic decrement.) We solve for the internal friction by substituting (31) into (30) and equating imaginary parts to find that

$$\frac{\Delta}{\pi} = \tan\phi \tag{32}$$

Because of the potential difficulties of measuring two successive amplitudes, we proceed a step further. Letting t_e denote the time required for the amplitude to decrease to $\frac{1}{e}$ of its original value, we note that

$$t_e = \frac{1}{f\Delta} = \frac{2\pi}{\omega\Delta} \tag{33}$$

Rearranging and substituting (33) into (32) then yields

$$\tan\phi = \frac{1}{\pi f t_e} = \frac{2}{\omega t_e} \tag{34}$$

Our second measure of internal friction is then given in terms of the time it takes the amplitude of vibration (and therefore of the sound) to decrease to some fraction of the original amplitude.

3.3 Specific Loss

We now have in hand two measures of internal friction (29) and (34). Examination of these relations reveals that $\tan\phi$ is a function of frequency. For our purposes, it would be more convenient to have a single characterization of anelastic behavior rather than an entire "damping spectrum". Such a measure is available in terms of loss per cycle, known as "specific loss". To motivate this measure we notice from (21) that for $(\omega\tau)^2 \gg 1$, $\omega\tan\phi$ is essentially independent of frequency. Empirical investigations of solids (Bennewitz & Rotger, 1936; Gemant & Jackson, 1937) confirm that this relation is true for a very wide range of vibrational frequencies and materials. Therefore, we define specific loss as

$$\delta = \omega\tan\phi \qquad\qquad (35)$$

as our measure of anelasticity and as our key to classifying solid materials by sound.

3.4 A Constraint Curve

The final task we face is to develop a method of telling when our two measures of internal friction will yield an accurate characterization of a given sample. For example, let us say an object is struck and our auditory system recognizes its pitch as $A_8 = 880$ Hz with a decay rate of 1 sec for t_e. By equation (34) we can recover a measure of internal friction, and hence characterize the material. But how do we know this answer is correct? To confirm that the material behaves as our standard anelastic linear solid, we need a second, corroborating measure. But of course, this can be obtained by recovering the bandwidth, Q^{-1}, of the sound about 880 Hz, using equation (29). In this case, if each measure does not yield the same value for $\tan\phi$, then we must disregard this characterization.

By plotting t_e versus $Q\omega^{-1}$ as shown in Figure 4, we can construct a constraint line that aolds for materials obeying our standard anelastic linear model. The line is independent of frequency because each measure of internal friction has been recast as "specific loss" (35). Clearly, all valid measurements must lie along this constraint curve. Here we show how particular material types spread out along the curve with metal at the upper right and rubber at the lower left. The data have been collected from Gemant & Jackson (1937), Berg & Stork (1982) and the *Handbook of Chemistry and Physics*. Exactly how fine a distinction can be realistically made along the constraint curve is largely an empirical question (Waller, 1938; Warren & Verbrugge, 1984).
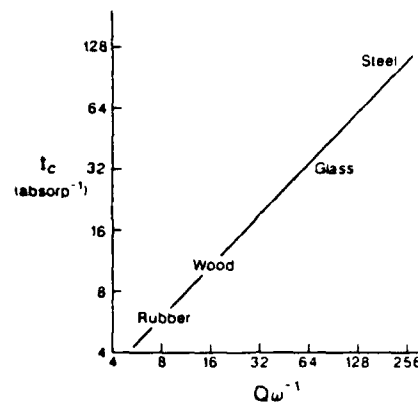
Figure 4 The constant curve relating the two measures of anelasticity for our standard linear solid model. Decay: t_c; tuning width: Q^{-1}.

4.0 Summary

We have described how to make auditory measures of an intrinsic parameter of solid materials. The parameter chosen is called specific loss, and is related to internal friction, a measure of anelasticity for a given material. The two measures presented are related to the width of the resonant peak of a sound and the time of its decay. Crucial to our development has been an understanding of a standard anelastic linear system, a physical model of the mechanics of solid vibration. We hope this work demonstrates the usefulness of using an explicit physical model of sound production in attempts to study sound recognition.

5.0 References

Bennewitz, V.K. & Rotger, H. (1936) Uber die Reibung fester Korper; Absorptions-frequenzen von Metallen im akustischen Gebiet, *Phys. Zeitschr.*, 37:578 – 588.

Berg, R.E. & Stork, D. (1982) *The Physics of Sound.* New Jersey: Prentice-Hall.

Brady, M., Hollerbach, J.M., Johnson, T.L., Lozano-Perez, T. & Mason, M.T. (1982) *Robot Motion: Planning and Control.* Cambridge, MA: MIT Press.

Horn, B.K.P. (1970) Shape from shading: a method for obtaining the shape of a smooth opaque object from one view. MIT Project MAC Internal Report TR-79 and MIT A.I. Lab. Technical Report 232.

Gemant, A. & Jackson, W. (1937) The measurement of internal friction in some solid materials, *Phil. Mag.*, 157:960–983.

Marr, D. (1982) *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: Freeman.

Nowick A.S. & Berry, B.S. (1972) *Anelastic Relaxation in Crystaline Solids.* New York: Academic Press.

Richards, W.A. (1988) *Selections in Natural Computation*, Cambridge, MA: MIT Press.

Waller, M.D. (1938) Magneto-damping in nickel, *Phys. Soc. Proc. (London)*, 50:144–146.

Warren, W.H. & Verbrugge, R.R. (1984) Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, 10:704–712.

Wert, C.A. (1986) Internal friction in solids, *J. Appl. Phys.*, 60(6):1888–1895.

Zener, C. (1948) *Elasticity and Anelasticity of Metals* Chicago, IL: University of Chicago Press.

References[1] and Publications[2]

Ballard, D.H. & Brown, C.M. (1982) *Computer Vision.* Prentice-Hall: Englewood Cliffs, N.J.

Binford, T.O. (1971) Visual perception by computer. Paper presented at IEEE Conference on Systems and Control, Miami, December.

Blinn, J.F. (1977) Models of light reflection for computer synthesized pictures. *Computer Graphics,* 11:192–198.

Blinn, J.F. (1978) Simulation of wrinkled surfaces. *SIGGRAPH '78, Proceedings, ACM, Computer Graphics,* pp. 286–292.

Blinn, J.F. (1982) Light reflection functions for simulation of clouds and dusty surfaces. *Computer Graphics,* 16:21–29.

Brodatz, P. (1966) *Textures: A Photographic Album for Artists and Designers.* Dover: New York.

Carpenter, L.C. (1980) "Vol Libre". *Computer Generated Movies.* (See also *ACM,* 25:371–384, 1982.)

Cook, R.L. & Torrance, K.E. (1982) A reflectance model for computer graphics. *ACM Trans. on Graphics,* 1:7–24.

Fournier, A. & Reeves, W.T. (1987) Special issue on modelling Natural phenomena. *ACM Trans. Graphics,* 6:165–166.

Grzywacz, N.M. & Hildreth, E. (1987) The incremental rigidity scheme for recovering structure from motion. *J. Opt. Soc. Am. A,* 3:503–518.

Heeger, D. & Pentland, A. (1987) Texture segmentation by use of fractal statistics. *Pat. & Mach. Intell.,* in press.

*Honig, D. & Richards, W. (1986) Why does wood look like wood? *J. Opt. Soc. Amer. A,* 2:P29, December.

Horn, B.K.P. (1977) Understanding image intensities. *Artif. Intell.,* 8:201–231.

Horn, B.K.P. (1987) *Robot Vision.* M.I.T. Press: Cambridge, MA.

Kajiya, J.T. & von Herzen, B.P. (1984) Ray tracing volume densities. *Computer Graphics,* 18:165–174 (SIGGRAPH Conference Proceedings).

Koenderink, J.J. & van Doorn, A.J. (1986) Depth and shape from differential perspective in the presence of bending deformations. *J. Opt. Soc. Am. A,* 3:242–249.

Kube, P. & Pentland, A. (1986) On the imaging of fractal surfaces. SRI Tech. Note 390, Artificial Intelligence Center, SRI, Menlo Park, CA.

*Kung, T.J. & Richards, W. (1986) Rendering water. *J. Opt. Soc. Amer. A,* 2:P29, December.

---

[1] For additional references, see end of individual sections.

[2] Asterisk indicates publication supported by the contract.

*Kung, T.J. & Richards, W. (1988) Infering "water" from images. Chapter 16 in *Natural Computation*, W. Richards (ed.), MIT Press: Cambridge, MA.

Lewis, J.P. (1987) Generalized stochastic subdivision. *ACM Trans. on Graphics*, 6:167–190.

Marr, D. (1982) *Vision: A Computational Investigatin into the Human Representation and Processing of Visual Information*. Freeman: San Francisco.

Pentland, A. (1983) Fractal-based description. *Proc. Int. Joint Conf. on Artificial Intell.*, 183, Karlsruhe, Germany, August.

Pentland, A. (1984) Fractal-based description of natural scenes. *Analysis & Mach. Intell.*, 6:661–675.

Pentland, A.P. (1988) Fractal-based descriptions of surfaces. Chapter 20 in *Natural Computation*, W. Richards (ed.), MIT Press: Cambridge, MA.

Phong, B.-T. (1975) Illumination for computer generated images. *Comm. ACM*, 18(6):311–317.

Richards, W. (1982) How to play twenty questions with nature. MIT A.I. Lab. Memo 660.

*Richards, W. (1985) Structure from Motion. *J. Opt. Soc. Amer. A*, 2:343–349.

*Richards, W. (1988) *Natural Computation*. MIT Press: Cambridge, MA.

Richards, W. & Bobick, A. (1988) Playing twenty questions with nature. In *Computational Processes in Human Vision*, Z. Pylyshyn (ed.), Ablex: Norwood, NJ.

*Richards, W. & Ullman, S. (1987) *Image Understanding 1985-86*, Ablex Publishing, Norwood, N.J.

Rubin, J. & Richards, W. (1982) Color vision and image intensities: when are changes material? *Biol. Cyber.*, 45:215–226.

*Rubin, J. & Richards, W. (1987) Spectral categorization of materials. In *Image Understanding 1985–86*, W. Richards & S. Ullman (eds.), Ablex: Norwood, NJ.

*Rubin, J. & Richards, W. (1988) The visual perception of moving parts. *Jrl. Opt. Soc. Amer. A*, 5:2045–2049.

Ullman, S. (1984) Maximizing rigidity: the incremental recovery of 3D structure from rigid and rubbery motion. *Perception*, 13:255–274.

*Ullman, S. (1986) The optical flow of planar surfaces. *Spatial Vision*, pp. 263–276. Also MIT A.I. Lab Memo 870.

*Ullman, S. & Yuille, A. (1987) Rigidity and smoothness of motion. MIT A.I. Lab. Memo 989.

Voss, R.F. (1981) Plate c13 in Mandelbrot's *Fractal Geometry of Nature*, Freeman: San Francisco.

*Wildes, R. & Richards, W. (1988) Recovering material properties from sound. Chapter 25 in *Natural Computation*, W. Richards (ed.), MIT Press: Cambridge, MA.